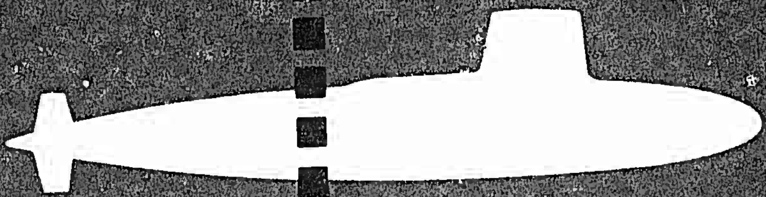
AD NUMBER
AD841587
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited. Document partially illegible.**
FROM: **Distribution authorized to U.S. Gov't. agencies only; Administrative/Operational Use; JUL 1968. Other requests shall be referred to Office of Naval Research, Code 462, Washington, DC 20360. Document partially illegible.**
AUTHORITY
onr ltr, 9 nov 1973

AD841537

SUBIC

[redacted]



Submarine Integrated Control

OFFICE OF NAVAL RESEARCH

GENERAL DYNAMICS CORPORATION
ELECTRIC BOAT DIVISION
GROTON, CONNECTICUT

DDC REPORT BIBLIOGRAPHY SEARCH CONTROL NO. /YATE4

AD-841 587 17/1
GENERAL DYNAMICS CORP GROTON CONN ELECTRIC BOAT DIV

PROCESSING OF DATA FROM SONAR SYSTEMS. VOLUME V
SUPPLEMENT 1. (U)

DESCRIPTIVE NOTE: ANNUAL REPT. 1 JUL 66-1 JUL 67,
JUL 68 206P CHANG,JOHN H. ;MCDONALD,
VERNE H. ;SCHULTHEISS,PETER M. ;TUTEUR,FRANZ
B. ;
REPT. NO. U417-68-079
CONTRACT: NONR-2512(00)

UNCLASSIFIED REPORT
~~DISTRIBUTION: USGO: OTHERS TO CHIEF, OFFICE OF NAVAL RESEARCH, ATTN: CODE 462, WASHINGTON, D. C. 20360.~~
SUPPLEMENTARY NOTE: PREPARED IN COOPERATION WITH YALE
UNIV., NEW HAVEN, CONN. DEPT. OF ENGINEERING
AND APPLIED SCIENCE. SEE ALSO AD-831 579.

DESCRIPTORS: (*SONAR SIGNALS, DATA PROCESSING
SYSTEMS), (*UNDERWATER OBJECTS, DETECTION),
SONAR ARRAYS, CLIPPER CIRCUITS, NOISE,
INTERFERENCE, UNDERWATER TRACKING, HYDROPHONES,
DELAY LINES, ELECTRIC FILTERS, CORRELATORS,
SIGNAL-TO-NOISE RATIO, DOPPLER EFFECT, UNDERWATER
SOUND, SOUND TRANSMISSION, SONAR TARGETS,
REPORTS (U)
IDENTIFIERS: SUBIC PROJECT,
*CLIPPING(ELECTRONICS), *SUBIC(SUBMARINE
INTEGRATED CONTROL), *SUBMARINE INTEGRATED
CONTROLS (U)

THIS DOCUMENT PURSUES THE GENERAL SUBJECT OF PASSIVE SONARS OPERATING IN AN ANISOTROPIC NOISE ENVIRONMENT, AND HAS TAKEN TWO NEW DIRECTIONS. ENVIRONMENTS CONTAINING NOT ONE BUT SEVERAL POINT SOURCES OF INTERFERENCE OR A SPATIALLY DISTRIBUTED INTERFERENCE ARE STUDIED. CONVENTIONAL AS WELL AS OPTIMAL DETECTORS WERE ANALYZED. THE SECOND DIRECTION WAS AN EXAMINATION OF THE EFFECT OF SINGLE PLANE WAVE INTERFERENCE ON TRACKING ACCURACY. THIS VOLUME ALSO CONTINUES THE STUDY OF ACTIVE SONAR SYSTEMS INITIATED IN VOLUME IV, AND INITIATES AN EFFORT TO DEAL WITH THE SIGNAL DETECTION AND EXTRACTION PROBLEM IN A NOISE ENVIRONMENT WHOSE STATISTICAL PROPERTIES ARE LARGELY OR WHOLLY UNKNOWN. (AUTHOR) (U)

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) General Dynamics Corporation Electric Boat division Groton, Connecticut	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

PROCESSING OF DATA FROM SONAR SYSTEMS, VOLUME V SUPPLEMENT 1

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Annual Report, 1 July 1966 to 1 July 1967

5. AUTHOR(S) (First name, middle initial, last name)

John H. Chang, Verne H. McDonald, Peter M. Schultheiss, and Franz B. Tuteur

6. REPORT DATE July 31, 1968	7a. TOTAL NO. OF PAGES 213	7b. NO. OF REFS ---
8a. CONTRACT OR GRANT NO. NOnr 2512(00) b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) U417-68-079 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

Each transmittal of this document outside the agencies of the U.S. Government must have approval of the Office of Naval Research.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Washington, D.C.

13. ABSTRACT

Volume V, and its supplement, further pursues the general subject of passive sonars operating in an anisotropic noise environment, and has taken two new directions. Environments containing not one but several point sources of interference or a spatially distributed interference are studied. Conventional as well as optimal detectors were analyzed. The second direction was an examination of the effect of single plane wave interference on tracking accuracy. This volume also continues the study of active sonar systems initiated in Volume IV, and initiates an effort to deal with the signal detection and extraction problem in a noise environment whose statistical properties are largely or wholly unknown.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

GENERAL DYNAMICS CORPORATION
Electric Boat division
Groton, Connecticut

PROCESSING OF DATA FROM SONAR SYSTEMS

VOLUME V

SUPPLEMENT 1

by

John H. Chang
Verne H. McDonald
Peter M. Schultheiss
Franz B. Tuteur

Yale University

Examined: J. W. Herring
SUBIC Project Manager

Approved: Dr. A. J. van Woerkom
Chief Scientist

U417-68-079
July 31, 1968

TABLE OF CONTENTS

FOREWORD

This is an unclassified supplement to Volume V of a series of reports describing work performed by Yale University under subcontract to Electric Boat division of General Dynamics Corporation. Volume V and this supplement cover the period 1 July 1966 to 1 July 1967. Electric Boat is prime contractor of the SUBIC (SUBmarine Integrated Control) Program under Office of Naval Research contract NOnr 2512(00). LCDR. E. W. Lull, USN, is Project Officer for ONR; J. W. Herring is Project Manager for Electric Boat division under the direction of Dr. A. J. van Woerkom.

Preceding page blank

THE EFFECT OF CLIPPING ON THE

PERFORMANCE OF REPLICA CORRELATORS

by

Peter M. Schultheiss

Progress Report No. 31

General Dynamics/Electric Boat Research

(8050-31-55001)

June 1967

DEPARTMENT OF ENGINEERING

AND APPLIED SCIENCE

YALE UNIVERSITY

Summary

This report deals with the effect of clipping on the performance of an active sonar system using conventional beamforming techniques followed by replica correlation. Detection as well as range and Doppler estimation are considered.

Two basic assumptions are made throughout the analysis:

a) The noise field (ambient or reverberation) is Gaussian and has the same power level at each hydrophone.

b) The input signal-to-noise ratio at each hydrophone is small.

In addition much, though not all, of the work assumes a transmitted signal narrow in bandwidth compared with its center frequency. The model for reverberation noise postulates a series of stationary, Poisson distributed scattering centers. The array geometry is quite arbitrary, but certain computations require beam patterns narrow enough to permit approximations of the form $\sin\theta \doteq \theta$ over the effective dimensions of the pattern.

The quantity of primary interest is the clipping loss R, defined as the output signal-to-noise ratio of the clipped instrumentation divided by the output signal-to-noise ratio of the unclipped (but otherwise identical) instrumentation.

The following results are obtained for detection:

1) If the noise (ambient or reverberation) is independent from hydrophone to hydrophone one can demonstrate with complete generality that $R \leq 1$. One can further demonstrate that, without additional restrictions on signal and noise, a lower bound of $R = 0$ can be approached arbitrarily closely. To set meaningful bounds on clipping loss one must therefore restrict the class of admissible signals and

noises. A practically realistic and analytically fruitful restriction is the assumption of narrow-band signals, which underlies all remaining results.

2) If the noise (ambient or reverberation) is independent from hydrophone to hydrophone and if signal and noise are confined to the same frequency band (narrow compared with the center frequency) the clipping loss is bounded by $0.89 \leq R \leq 1$.

3) If the noise does not fall into the same frequency band as the signal, large clipping losses can occur. This is practically important in a reverberation limited environment when the target return is subject to large Doppler shifts. In such situations R can approach arbitrarily close to zero if the Doppler shift is large enough.

4) If statistical dependences are allowed between the noise at different hydrophones, one requires further restrictions before useful lower bounds can be set on R . An example is worked out to demonstrate that, even with purely isotropic ambient noise, values of R appreciably below 0.89 may be obtained. However, the example requires such careful matching of array geometry with carrier wavelength that it appears to fall more into the category of analytical pathologies than into that of practically important situations. A search (by no means exhaustive) for more realistic examples in which ambient noise would produce values of R below 0.89 led to negative results.

5) Probably the most important situation from a practical point of view is that of a reverberation limited environment. It was

therefore studied in some detail. The most useful results were obtained under the assumption that the array dimensions are small compared with the wavelength of the highest modulating frequency, (i.e., the wavelength of the maximum frequency deviation from the carrier). The effect of this assumption is to permit complete separation of spatial and temporal effects in the reverberation. In the absence of target Doppler shifts one can then establish with considerable generality that $R \geq 0.89$. In the presence of target Doppler shifts one has, of course, the phenomenon discussed in 3). The key assumption concerning array dimensions can be weakened considerably if the beam pattern is narrow.

When one considers range and Doppler measurements one finds, not surprisingly, that the exact clipping loss depends to some extent on the specific instrumentation. It is therefore not possible to draw conclusions of quite the same generality as in the analysis of detection. Range (Doppler shift) is measured by cross-correlating the target return with a replica of the transmitted signal and "locating" the resulting correlation function in time (frequency). Different instrumentations result from different functional definitions of the term "location". It appears reasonable to speculate - and several sample computations tend to confirm this - that most reasonable definitions of "location" would lead to rather similar instrumentations, and in particular to instrumentations with very similar sensitivity to clipping. This report concerns itself primarily with range (Doppler) measurement in a reverberation limited environment. Range (Doppler shift) is measured by comparing

cross-correlations with two replicas of slightly different delay (frequency). Doppler shift is assumed known during the range measurement and range during the Doppler measurement. Clipping loss is defined as the ratio of the rms range (Doppler) errors with and without clipping. The results are

6) Under the conditions specified in 5) the clipping loss factor for range measurement has a lower bound not significantly different from the figure of 0.89 obtained for detection.

7) As might be anticipated from 3), the clipping loss in Doppler measurement depends heavily on the target Doppler shift. Separating out this effect by working with zero target Doppler shift, one can again show that the clipping loss factor has a lower bound close to 0.89.

Combining all of the above, one is lead to the following general conclusion: Serious clipping losses arise in a reverberation limited environment when the target Doppler shift is large enough to move the target return largely out of the reverberation band. In most other practically interesting situations the clipping loss in detection, ranging and Doppler estimation is limited to a factor of the order of 0.89, equivalent to about 1 db of input signal-to-noise ratio.

I. Introduction

This report is concerned with the effect of clipping on the performance of an active sonar array. Correlation with a replica (delayed and Doppler shifted) of the transmitted signal is used as the basic signal processing technique.

The general block diagram for detection is shown in Figure 1. The array geometry is entirely general.

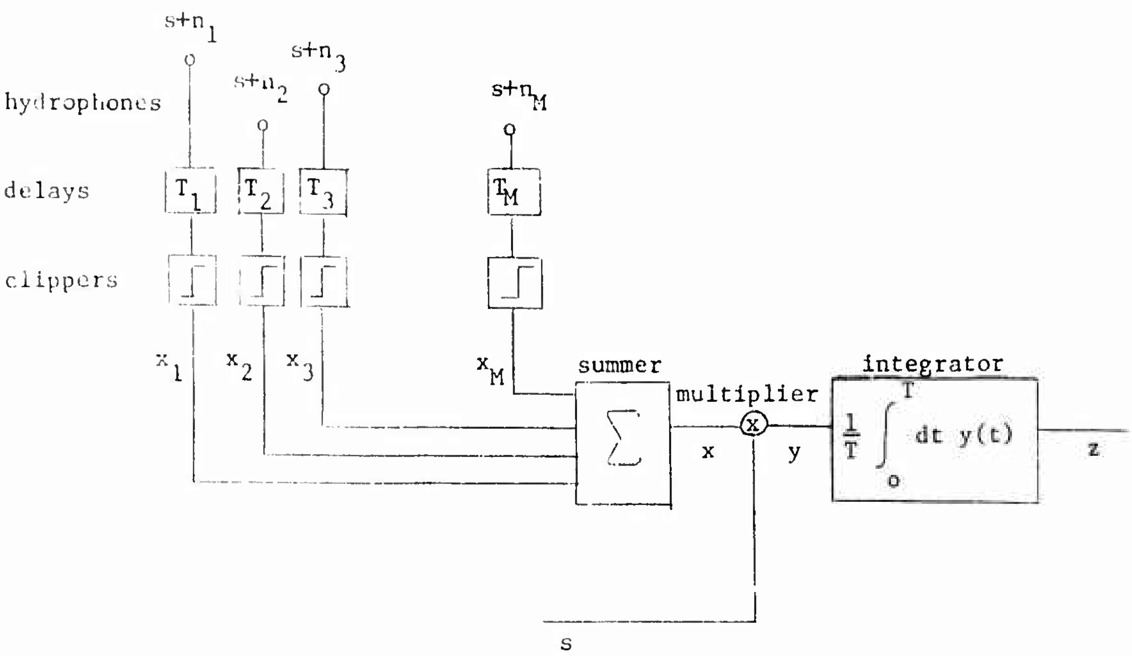


Figure 1

The delays τ_i bring the signal components of all hydrophone outputs into alignment. The clippers[1] then generate a set of signals $\{x_i(t)\}$ given

---

[1] In practice the clippers would probably precede the delays (which would then be digital). It is clear from physical reasoning that this interchange would not alter the $x_i(t)$.

by

$$x_i(t) = \mathrm{sgn}\left[s(t)+n_i(t-\tau_i)\right] = \begin{cases} 1 & \text{if} \quad s(t)+n_i(t-\tau_i) > 0 \\ 0 & \text{if} \quad s(t)-n_i(t-\tau_i) < 0 \end{cases} \tag{1}$$

The $x_i(t)$ are summed and multiplied by a suitable replica of $s(t)$. The resulting $y(t)$ is finally smoothed over a period T comparable to the duration of the signal. If the noise has zero mean the average output $\bar{z}$ is zero in the absence of a signal component at the hydrophones. When a signal component is present at the hydrophones $\bar{z}$ will differ from zero. Hence, if the output $z(T)$ at time T exceeds a preset threshold (depending on signal-to-noise ratio and allowed false alarm rate) one concludes that a target is indeed present.

To measure range (or Doppler) one might employ two replicas with slightly different delays (or Doppler shifts), multiply each by the clipped and summed hydrophone outputs (x) and use the difference between the resulting y's as a measure of range (or Doppler shift). A possible implementation is discussed in somewhat more detail at a later point.

The noise field is assumed to be Gaussian. This is reasonable even in a reverberation limited environment as long as no major portion of the noise power is contributed by large scatterers (false targets). [See Report No. 27].

II. General Relations for Detection

The output $y(t)$ of the multiplier can be written in the form

$$y(t) = \sum_{i=1}^{M} x_i(t)\ s(t) = \sum_{i=1}^{M} \mathrm{sgn}\left[s(t)+n_i(t-\tau_i)\right]\ s(t) \tag{2}$$

The delay of the replica is here assumed to be perfectly matched to the signal delay.

The mean value of the detector output is

$$\bar{z} = \sum_{i=1}^{M} \frac{1}{T} \int_0^T s(t)\ \overline{\mathrm{sgn}\left[s(t)+n_i(t-\tau_i)\right]}\ dt \tag{3}$$

Now

$$\overline{\mathrm{sgn}\left[s(t)+n_i(t-\tau_i)\right]} = \Pr\{\left[s(t)+n_i(t-\tau_i)\right] > 0\} - \Pr\{\left[s(t)+n_i(t-\tau_i)\right] < 0\} \tag{4}$$

For zero-mean Gaussian $n_i(t)$

$$\Pr\{\left[s(t)+n_i(t-\tau_i)\right] > 0\} = \frac{1}{\sqrt{2\pi N}} \int_0^\infty e^{-\frac{\left[n-s(t)\right]^2}{2N}}\ dn = \tfrac{1}{2}\left[1+\mathrm{erf}\ \frac{s(t)}{\sqrt{2N}}\right] \tag{5}$$

where
$$\mathrm{erf}\ v = \frac{2}{\sqrt{\pi}} \int_0^y e^{-x^2}\ dx \tag{6}$$

and N is the average noise power.

Similarly

$$\Pr\{\left[s(t)+n_i(t-\tau_i)\right] < 0\} = \frac{1}{\sqrt{2\pi N}} \int_{-\infty}^0 e^{-\frac{\left[n-s(t)\right]^2}{2N}}\ dn = \tfrac{1}{2}\left[1-\mathrm{erf}\ \frac{s(t)}{\sqrt{2N}}\right] \tag{7}$$

It follows that

$$\overline{\operatorname{sgn}\left[s(t)+n_i(t-\tau_i)\right]} = \operatorname{erf} \frac{s(t)}{\sqrt{2N}} \tag{8}$$

when $\max\limits_{t} \; s(t) << \sqrt{2N}$

$$\overline{\operatorname{sgn}\left[s(t)+n_i(t-\tau_i)\right]} \cong \sqrt{\frac{2}{\pi}} \; \frac{s(t)}{\sqrt{N}} \tag{9}$$

Hence $$\bar{z} \cong \frac{\sqrt{2}M}{\sqrt{\pi}T\sqrt{N}} \int_0^T dt \; s^2(t) \qquad \text{for} \quad \max_{t} \; s(t) << \sqrt{2N} \tag{10}$$

For a figure of merit of the detector we choose as usual the output signal-to-noise ratio, i.e., the average output due to the signal divided by the rms value of the output fluctuation. If the input signal-to-noise ratio is low, the output fluctuation is very largely due to the noise component of the input, so that

$$\overline{z^2(t)} = \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{1}{T^2} \int_0^T dt \int_0^T d\lambda \; s(t) \; s(\lambda) \; \overline{\operatorname{sgn}\left[s(t)+n_i(t-\tau_i)\right] \; \operatorname{sgn}\left[s(\lambda)+n_j(\lambda-\tau_j)\right]}$$

$$\cong \frac{1}{T^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \int_0^T dt \; s(t) \int_0^T d\lambda \; s(\lambda) \; \overline{\operatorname{sgn}\left[n_i(t-\tau_i)\right] \; \operatorname{sgn}\left[n_j(\lambda-\tau_j)\right]} \tag{11}$$

From a well-known result in noise theory

$$\overline{\operatorname{sgn}\left[n_i(t)\right] \; \operatorname{sgn}\left[n_j(\lambda)\right]} = \frac{2}{\pi} \sin^{-1} \left[\rho_{ij}(t-\lambda)\right] \tag{12}$$

where $\rho_{ij}(\tau)$ is the normalized cross-correlation function of the noise received at the i^{th} and j^{th} hydrophones. Hence, for low input signal-

to-noise ratio

$$\overline{z^2(t)} = \frac{1}{T^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \frac{2}{\pi} \sin^{-1}\left[\rho_{ij}(t-\lambda-\tau_i+\tau_j)\right] \tag{13}$$

From Equations (10) and (13) the desired output signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_{0 \text{ clipped}} = \frac{\sqrt{\frac{2}{\pi}}\ \frac{M}{T\sqrt{N}} \int_0^T s^2(t)\ dt}{\sqrt{\frac{1}{T^2} \int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{2}{\pi} \sin^{-1} \rho_{ij}(t-\lambda-\tau_i+\tau_j)}} \tag{14}$$

In the absence of clipping

$$\overline{z} = \frac{M}{T} \int_0^T s^2(t)\ dt \tag{15}$$

and

$$\overline{z^2} = \frac{1}{T^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ N_{ij}\ \rho_{ij}(t-\lambda-\tau_i+\tau_j) \tag{16}$$

where

$$N_{ij}\ \rho_{ij}(t-\lambda) = \overline{n_i(t)\ n_j(\lambda)} \tag{17}$$

Hence, the output signal-to-noise ratio in the unclipped case is

$$\left(\frac{S}{N}\right)_{0 \text{ unclipped}} = \frac{\frac{M}{T} \int_0^T s^2(t)\ dt}{\sqrt{\frac{1}{T^2} \int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} N_{ij}\ \rho_{ij}(t-\lambda-\tau_i+\tau_j)}} \tag{18}$$

The present study is concerned with performance degradation due to clipping. Hence the ratio R of Equations (14) and (18) serves as a convenient criterion.

$$R = \frac{(S/N)_0 \text{ clipped}}{(S/N)_0 \text{ unclipped}} = \sqrt{\frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} N_{ij}\, \rho_{ij}(t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} N \sin^{-1}\left[\rho_{ij}(t-\lambda-\tau_i+\tau_j)\right]}} \tag{19}$$

If the average noise power is the same at each hydrophone, then $N_{ij} = N$ and Equation (19) reduces to

$$R = \sqrt{\frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}(t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \sin^{-1}\left[\rho_{ij}(t-\lambda-\tau_i+\tau_j)\right]}} \tag{20}$$

III. Noise Independent from Hydrophone to Hydrophone

An important special case is that of noise independent from hydrophone to hydrophone. In that situation, Equation (20) reduces to

$$R = \sqrt{\frac{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \rho(t-\lambda)}{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \sin^{-1}\left[\rho(t-\lambda)\right]}} \tag{21}$$

where $\rho(\tau) = \rho_{ii}(\tau)$, $i = 1, 2 \ldots M$, is the normalized autocorrelation function of the noise at each hydrophone.

It is a simple matter to show that Equation (21) can never exceed unity. For greater ease in manipulation, consider the quantity $1/R^2$ and expand the inverse sine into a power series (convergent for all values of t and λ since $|\rho(t-\lambda)| \leq 1$) .

$$\frac{1}{R^2} = \frac{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \left[\rho(t-\lambda)+1/2\times 1/3\ \rho^3(t-\lambda)+1/2\times 3/4\times 1/5\ \rho^5(t-\lambda)+\ldots\right]}{\int_0^\infty dt\ s(t) \int_0^\infty d\lambda\ s(\lambda)\ \rho(t-\lambda)}$$

$$= 1 + 1/6\ \frac{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \rho^3(t-\lambda)}{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \rho(t-\lambda)} + \frac{3}{40}\ \frac{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \rho^5(t-\lambda)}{\int_0^T dt\ s(t) \int_0^T d\lambda\ s(\lambda)\ \rho(t-\lambda)} + \ldots \tag{22}$$

$\rho(\tau)$, being a correlation function, is positive definite. Hence, the

denominator in each term is positive, regardless of the form of $s(t)$. Furthermore, $\rho(\tau)$ is Fourier inverse of a non-negative spectral function $G(w)$. It follows that $\rho^n(\tau)$ is the Fourier inverse of the n-fold convolution of $G(w)$ with itself. This convolution is clearly non-negative, so that $\rho^n(\tau)$ is positive definite. Hence, all the numerators in Equation (22) are non-negative. Thus all terms of the equation are non-negative and one concludes

$$\frac{1}{R^2} \geq 1$$

or

$$R^2 \leq 1 \tag{23}$$

In order to set a lower bound on R it is necessary to make subsidiary assumptions. In the absence of any restrictions on signal and noise properties one might postulate a signal confined to one frequency band and a noise confined to a different, disjoint frequency band. In such an environment the unclipped detector would be essentially perfect,[1] while the clipping process could shift appreciable noise power into the signal band. Thus values of R arbitrarily close to zero could be obtained in principle by sufficiently artificial choices of signal and noise spectra. Reasoning more formally, one can proceed as follows to establish the impossibility of finding a general lower bound on R in excess of zero:

Replace the integrals in Equation (22) by sums. The implied sampling in time can be very rapid, so that the approximation is arbitrarily good.

---

[1]Limited only by the impossibility of maintaining perfectly disjoint spectra with a signal of finite duration.

A typical term in Equation (22) can then be written in the form

$$A_n = k_n \frac{\sum_i \sum_j a^n_{ij} s_i s_j}{\sum_i \sum_j a_{ij} s_i s_j} \tag{24}$$

where

$$k_n = \frac{1 \times 3 \times 5 ---- (n-2)}{2 \times 4 \times 6 ---- (n-1)} \times \frac{1}{n} \tag{25}$$

and

$$a_{ij} = \rho(t_i - \lambda_j) \; ; \qquad s_i = s(t_i)$$

Then

$$A_n = k_n \frac{\sum_i \sum_j \left[(a^n_{ij} - a_{ij}) + a_{ij}\right] s_i s_j}{\sum_i \sum_j a_{ij} s_i s_j}$$

$$= k_n \left\{ 1 - \frac{\sum_i \sum_j (a_{ij} - a^n_{ij}) s_i s_j}{\sum_i \sum_j a_{ij} s_i s_j} \right\} \tag{26}$$

Now choose

$$s_i = \begin{cases} 1 & \text{for } i = \ell \\ -1 & \text{for } i = k \\ 0 & \text{all other } i \end{cases} \tag{27}$$

Then [illegible] $a^n_{ii} - a_{ii} = 0$ and one obtains for the chosen s_i

$$A_n = E_n \left(1 + \frac{a_{k\ell}}{1} [illegible] \frac{a_{k\ell}^{n}}{a_{\ell}} [illegible] \right)$$

$$= [illegible]_n [illegible] (a_{k\ell})^{1} \qquad (28)$$

It follows that

$$[illegible] \frac{[illegible] (n + 2)}{[illegible] 1)} \qquad (29)$$

This upper [illegible] by choosing the times t_k and [illegible] sufficiently close together. With A_n at this upper bound, Equation [illegible] reads

$$\frac{1}{R^2} = 1 + [illegible] \frac{1 \cdot [illegible]}{2 [illegible]} + \frac{1 \cdot 3 [illegible]}{2 \cdot 4 [illegible]} + \ldots$$

$$= [illegible] x) [illegible] \lim_{x \to 1} \frac{1}{\sqrt{1 - x^2}} = \infty \qquad (30)$$

Thus, for any fixed [illegible] it is possible to find a signal $s(t)$ which causes R to be arbitrarily close to zero.

Most practical [illegible] are very different from the cases just [illegible] active sonar signal would almost certainly be of the form

$$[illegible] \qquad (31)$$

where $s_1(t)$ is [illegible] determining the pulse shape) and [illegible] modulation. The bandwidths of $s_1(t)$ [illegible] regarded as small compared with ω_0 [illegible]

Given a narrowband signal, it appears reasonable to assume that the noise is a narrowband process also. For if the noise is primarily reverberation, its spectral properties are determined by the signal, while in an ambient noise limited environment one would employ filters matched to the bandwidth of the signal in order to improve signal-to-noise ratio. If there is no Doppler shift in the target signal the center frequency of the noise will be w_0.[1] Then $\rho(\tau)$ would assume the form

$$\rho(\tau) = \rho_1(\tau) \cos w_0\tau \tag{32}$$

The bandwidth of $\rho_1(\tau)$ is small compared with w_0. Clearly $\rho_1(0) = 1$ and $|\rho_1(\tau)| \le 1$ for all τ_1. Typical forms of $\rho_1(\tau)$ might be $\rho_1(\tau) = e^{-a|\tau|}$ and $\rho_1(\tau) = e^{-a\tau^2}$ both of which have the computationally useful property $\rho_1(\tau) \geq 0$ for all τ. This property will be assumed in the immediately following computations.

Using Equations (31), (32) and (25) a typical term of Equation (22) can now be written in the form

$$A_n = k_n \frac{\int_0^T dt \int_0^T d\lambda\, s_1(t)\, s_1(\lambda) \cos[w_0 t+\phi(t)] \cos[w_0\lambda+\phi(\lambda)]\, \rho_1^n(t-\lambda) \cos^n w_0(t-\lambda)}{\int_0^T dt \int_0^T d\lambda\, s_1(t)\, s_1(\lambda) \cos[w_0 t+\phi(t)] \cos[w_0\lambda+\phi(\sigma)]\, \rho_1(t-\lambda) \cos w_0(t-\lambda)} \tag{33}$$

---

[1] This would remain approximately true for targets moving sufficiently slowly relative to the receiver so that the Doppler shift is small compared with the signal bandwidth.

Since n is an odd integer $\cos^n w_0(t-\lambda)$ can be expanded as follows

$$\cos^n w_0(t-\lambda) = C_{1n} \cos w_0(t-\lambda) + C_{3n} \cos 3w_0(t-\lambda) + \text{----} + C_{nn} \cos nw_0(t-\lambda) \qquad (34)$$

The constants C_{1n} [illegible] related to the binomial coefficients. In particular, a straightforward [illegible] yields

$$C_{1n} = \frac{n!}{[illegible]} \text{ [illegible] } \frac{n}{n+1} \qquad (35)$$

Substituting [illegible] into Equation (33) and invoking the narrowband assumption [illegible] can use the Riemann-Lebesgue lemma to argue that only the term $C_{1n} \cos w_0(t-\lambda)$ of Equation (34) contributes significantly to the integral. Hence, Equation (33) becomes

$$A_n \cong K_n C_{1n} \frac{\int_0^T dt \int_0^T d\lambda\, s_1(t) [illegible] \cos\left[w_0 t+\phi(t)\right] \cos\left[w_0\lambda+\phi(\lambda)\right] \cos w_0(t-\lambda)}{\int_0^T dt \int_0^T d\lambda\, s_1(t) [illegible]_1(t-\lambda) \cos\left[w_0 t+\phi(t)\right] \cos\left[w_0\lambda+\phi(\lambda)\right] \cos w_0(t-\lambda)} \qquad (36)$$

Straightforward use of [illegible] yields

$$\cos\left[w_0 t+\phi(t)\right] \cos\left[w_0\lambda+\phi(\lambda)\right] \cos w_0(t-\lambda)$$

$$= \frac{1}{4} \cos\left[\phi(t)-\phi(\lambda)\right] + \text{[illegible]} + \frac{1}{4} \cos\left[2w_0 t+\phi(t)+\phi(\lambda)\right]$$

$$+ \frac{1}{4} \cos\left[2w_0\lambda+\phi(t)+\phi(\lambda)\right] \qquad (37)$$

Once more the Riemann-Lebesgue [illegible] employed to eliminate all but [illegible] (36) becomes

$$\Lambda_n = K C_n \ln \frac{\int_0^T dt \int_0^T d\lambda \; s_1(t) s_1(\lambda) \rho_1^n(t-\lambda) \; \cos[\phi(t)-\phi(\lambda)]}{\int_0^T dt \int_0^T d\lambda \; s_1(t) s_1(\lambda) \rho_1(t-\lambda) \; \cos[\phi(t)-\phi(\lambda)]} \tag{38}$$

All factors in the integrand except for the cosine terms are non-negative. As for the cosine terms note that $\rho_1(t-\lambda)$ becomes small for $(t-\lambda)$ larger than the correlation time of the noise process. For reverberation this correlation time is at most (stationary scatterers) equal to the correlation time of the signal, which in turn is the smaller of the following two quantities: 1) the signal duration, 2) the time within the signal pulse during which the modulation $\phi(t)$ changes by, roughly, one radian. Thus the cosine terms remain positive throughout the effective region of integration.[1] In the ambient noise limited case the noise spectrum would generally be shaped by bandlimiting filters (or possibly transducer

---

[1] The case of reverberation from stationary scatterers is treated with much greater rigor and generality in Section IV.

characteristic [illegible] and hence have a bandwidth at least equal to that of the signal. Its correlation time would therefore be no longer than that of the signal. [illegible] the above [illegible] remains valid.

Under these conditions, the integrands in both numerator and denominator of Equation [illegible] are positive over the effective range of integration. Since $\rho_1^n(t)$ [illegible]

$$A_n \; [illegible] \; \left[\frac{1 \cdot 3 \cdot \cdots (n-2)}{[illegible] \cdots (n-1)}\right]^2 \frac{1}{n+1} \qquad (39)$$

Hence, from Equation [illegible]

$$\frac{1}{R^2} \leq 1 + 2 \sum_{\substack{[illegible] \\ \text{odd } n}} \left|\frac{1 \cdot 3 \cdot 5 \cdot \cdots \cdot (n-2)}{2 \cdot 4 \cdot 6 \cdot \cdots \cdot (n-1)}\right|^2 \frac{1}{n+1} \qquad (40)$$

From the well-known inequality

$$\left[\frac{1 \cdot 3 \cdot 5 \cdots (\,[illegible]\,-1)}{2 \cdot 4 \cdot 6 \cdots (\,[illegible]\,)}\right]^2 \frac{1}{[illegible]\,n+2} \; [illegible] \; \frac{2}{\pi} \qquad (41)[1]$$

one obtains, upon setting [illegible]

$$[illegible] \; \frac{2}{\pi} \frac{1}{n^2} \qquad (42)$$

It is now a simple matter to bound $1/R^2$ from above by evaluating the first few terms of Equation (40) computationally and using Equation (42) to set upper bounds on the remaining terms. Carrying the exact computation from Equation (40) to [illegible] one obtains

---

[1] [illegible] AMIE.

$$\frac{1}{R^2} \leq 1.277 \tag{43}$$ [1]

Combining Equations (23) and (43) one can therefore rather generally bound the clipping loss for noise independent from hydrophone to hydrophone by

$$0.89 \leq R \leq 1 \tag{44}$$

One situation in which the assumptions leading to Equation (44) are not satisfied is that of an ambient noise dominated environment containing a strong narrowband noise component. Here the effective range of integration is determined by the correlation time of the signal component of Equation (38). It is clear from intuitive considerations that $\cos[\phi(t)-\phi(\lambda)]$ does not change sign over this interval. More specifically, consider a signal of the form

$$s(t) = e^{-\frac{t^2}{\tau_1^2}} \cos\left(w_0 t + \frac{k}{2} t^2\right) \tag{45}$$

Substitution into Equation (38) yields, after an elementary computation,

$$A_n = k_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dx\, e^{-\left(\frac{1}{2\sigma_T^2} + \frac{k^2 \sigma_T^2}{8}\right)x^2} \rho_1^n(x)}{\displaystyle\int_{-\infty}^{\infty} dx\, e^{-\left(\frac{1}{2\sigma_T^2} + \frac{k^2 \sigma_T^2}{8}\right)x^2} \rho_1(x)} \tag{46}$$ [2]

---

[1] The relation $1 + 1/9 + 1/25 + 1/49 + \cdots = \pi^2/8$ has been used in summing the infinite series above $n = 9$.

[2] The limits were extended from $[0, T_1]$ to $(-\infty, \infty)$ because in practice the period of integration would undoubtedly cover the effective duration of the pulse.

Since both integrands of Equation (46) are non-negative and $\rho_1^n(x) \leq \rho_1(x)$, Equation (39) [and hence Equation (44)] remains valid.

Thus it appears that the only practically important exception to Equation (44) [for noise independent from hydrophone to hydrophone] occurs when a rapidly moving target is being detected in a reverberation limited environment. Here the returned signal might be of the form

$$s(t) = s_1(t) \cos\left[(w_D + w_0) t + \phi(t)\right] \tag{47}$$

where w_D is the Doppler shift caused by the moving target.[1] Following the same procedure as in Equations (33) to (38) one arrives at the following equivalent of Equation (38)

$$A_n = K_n C_{1n} \frac{\int_0^T dt \int_0^T d\lambda \; s_1(t)\, s_1(\lambda)\, \rho_1^n(t-\lambda) \cos\left[w_D(t-\lambda) + \phi(t) - \phi(\lambda)\right]}{\int_0^T dt \int_0^T d\lambda \; s_1(t)\, s_1(\lambda)\, \rho_1(t-\lambda) \cos\left[w_D(t-\lambda) + \phi(t) - \phi(\lambda)\right]} \tag{48}$$

As a specific example, consider

$$s_1(t) = e^{-\frac{t^2}{2T^2}} \quad , \quad \phi(t) = \frac{K}{2} t^2 \quad , \tag{49}$$

a linearly frequency modulated pulse.

It is a simple matter to demonstrate [see Equation (88) with $i = j$, $d_{ii} = 0$] that $\rho_1(\tau)$ [for reverberation from stationary scatterers]

---

[1] The signal is assumed to be sufficiently narrowband so that the Doppler shift may be regarded as constant throughout the band.

assumes the form

$$\rho_1(\tau) = e^{-\left(\frac{1}{2\sigma_T^2} + \frac{K^2\sigma_T^2}{8}\right)\tau^2} \tag{50}$$

$\left(\frac{1}{2\sigma_T^2} + \frac{K^2\sigma_T^2}{8}\right)$ may be interpreted as the bandwidth of the signal and hence of the reverberation.

Substituting Equations (49) and (50) into Equation (48) and extending the ranges of integration to $(-\infty, \infty)$, one obtains, after some algebraic manipulation

$$A_n = K_n C_{1n} \exp\left\{ w_D \frac{n-1}{n+1} \frac{1}{\frac{4}{\sigma_T^2} + K^2\sigma_T^2} \right\} \tag{51}$$

If the signal bandwidth is determined primarily by the frequency modulation $K^2\sigma_T^2 >> \frac{4}{\sigma_T^2}$ and Equation (51) becomes

$$A_n \cong K_n C_{1n} \exp\left\{ w_D \frac{n-1}{n+1} K^2\sigma_T^2 \right\} \tag{52}$$

Conversely, if $K^2\sigma_T^2 << \frac{4}{\sigma_T^2}$, [little or no frequency modulation]

$$A_n \cong K_n C_{1n} \exp\left\{ \frac{n-1}{4(n+1)} \sigma_T^2 w_D^2 \right\} \tag{53}$$

In either case the coefficient A_n becomes large when w_D greatly exceeds the bandwidth of the signal. Under the same conditions the factor R will therefore become very much smaller than unity. It may, in fact, come arbitrarily close to zero if the Doppler shift is large enough compared

with the signal bandwidth. This is quite reasonable from a physical point of view: Under the postulated conditions the signal and reverberation spectra are essentially disjoint so that the unclipped detector operates in an environment that is almost noise-free in the signal band. Clipping, on the other hand, shifts some of the reverberation power into the signal band and therefore sharply degrades detector performance.

IV. Noise Dependent from Hydrophone to Hydrophone

When the noise field exhibits significant coherence from hydrophone to hydrophone it appears to be difficult to obtain results of the same generality as in the previous section. Thus even in the absence of Doppler shifts and with constraints on signal and noise, such as those specified by Equations (31) and (32), one can readily specify realistic spatial covariances for the noise which result in values of R lower than 0.89 [Equation (44)] by modest amounts. An example of this type is worked out in Appendix A, where an R of 0.74 is shown to be attainable, even for isotropic ambient noise. With sufficient ingenuity, one suspects, one could devise noise models yielding even lower values of R. However, even the simple example of Appendix A requires fairly special assumptions and one feels that the search for more extreme cases would lead to more and more artificial assumptions. It appears more rewarding, therefore, to abandon the search for extremes and turn to the question whether serious clipping loss is likely to occur in cases commonly encountered in practice.

Some qualitative insight into this question may be gained by restating Equation (20) in the frequency domain. Defining

$$G_{ij}(w) = \frac{1}{\pi} \int_{-\infty}^{\infty} \rho_{ij}(\tau)\, e^{-jw\tau}\, d\tau \tag{54}$$

and

$$P_{ij}(w) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{2}{\pi} \sin^{-1}\left[\rho_{ij}(\tau)\right] e^{-jw\tau}\, d\tau \tag{55}$$

one obtains the Fourier inverses

$$\rho_{ij}(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} G_{ij}(w)\, e^{jw\tau}\, dw \tag{56}$$

and

$$\frac{2}{\pi} \sin^{-1}\left[\rho_{ij}(\tau)\right] = \frac{1}{2} \int_{-\infty}^{\infty} P_{ij}(w)\, e^{jw\tau}\, dw \tag{57}$$

Substitution of Equations (56) and (57) into Equation (20) yields (after a few steps of computation)

$$R = \sqrt{\frac{2}{\pi}} \sqrt{\frac{\int_{-\infty}^{\infty} dw |S(w)|^2 \sum_{i=1}^{M} \sum_{j=1}^{M} G_{ij}(w)\, e^{jw(\tau_j - \tau_i)}}{\int_{-\infty}^{\infty} dw |S(w)|^2 \sum_{i=1}^{M} \sum_{j=1}^{M} P_{ij}(w)\, e^{jw(\tau_j - \tau_i)}}} \tag{58}$$

where

$$S(w) = \int_{0}^{T} dt\, s(t)\, e^{-jwt} \tag{59}$$

Since the observation interval $(0,T)$ would generally cover the entire signal pulse, $S(w)$ is in effect the Fourier transform of the signal.

The expression $\sum_{i=1}^{M} \sum_{j=1}^{M} G_{ij}(w)\, e^{jw(\tau_j - \tau_i)}$ in the numerator of Equation (58) is nothing other than the normalized noise power spectrum at point x in Figure 1 with the clippers removed. $\sum_{i=1}^{M} \sum_{j=1}^{M} P_{ij}(w) e^{jw(\tau_j - \tau_i)}$ is the normalized noise spectrum at the same point in the presence of clipping. Thus the numerator (denominator) integral in Equation (58)

represents the power outputs of a filter matched to the signal whose inputs is x in the unclipped (clipped) instrumentation. Since clipping tends to spread the spectrum, one would expect a smaller percentage of the power to fall within the filter band in the clipped than in the unclipped case. Hence R should generally tend to exceed $\sqrt{2/\pi} = 0.8$. Important exceptions to this rule would be expected in two cases:

1) The unclipped noise spectrum is centered at a frequency quite different from the center frequency of $|S(w)|^2$. This would be the case when the target return is subject to a strong Doppler shift. Reverberation noise is not subject to this shift and it is only through the clipping operation that a significant amount of noise power is transferred into the signal band.

2) Strong negative correlation between closely adjacent x_i in Figure 1 causes the terms $i \neq j$ in Equation (58) to subtract substantially from the power input into the filter matched to the signal. This effect can result in values of R smaller than $\sqrt{2/\pi}$ only if clipping reduces the negative correlation so that the effect is less pronounced in the denominator than in the numerator of Equation (58). To see when this might be the case, consider the inverse sine in Equation (20) expanded into a power series, as in III. The equivalent of Equation (22) is now

$$\frac{1}{K^2} = 1 + \frac{1}{6}\,\frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}^3(t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}(t-\lambda-\tau_i+\tau_j)}$$

$$+ \frac{3}{40}\,\frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}^5(t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}(t-\lambda-\tau_i+\tau_j)} \cdots \qquad (60)$$

Unless the peak negative value of $\rho_{ij}(\tau)$ is fairly close to unity for a significant number of $i \neq j$, the required subtraction in the denominator does not take place. On the other hand, if $\rho_{ij}(\tau)$ comes very close to (-1), $\rho_{ij}^3(\tau)$ is not too far from (-1) and a similar subtraction takes place in the numerator. Thus one looks for maximum clipping loss in cases where the negative correlation between closely adjacent phones is strong, but not strong enough so that $\rho_{ij}^3(\tau)|_{max}$ has a magnitude comparable to unity. Since these conditions on the ρ_{ij} are quite restrictive one is not surprised to find only modest decreases in R for even rather carefully constructed examples, such as the situation analyzed in Appendix A.[1]

---

[1]Note that the rather rapidly converging sequence of coefficients in Equation (60) demands integral ratios of at least the order of 5 before $1/K^2$ begins to increase very substantially.

To complete the discussion the clipping loss will now be computed for a fairly general case of operation in a reverberation limited environment. The most important restriction is the assumption of neglible Doppler shift, both in the target return and in the reverberation.[1]

The key step is the computation of the cross-correlation of the reverberation at the i^{th} and j^{th} hydrophone. If the signal assumes the narrow-band form of Equation (31) the reverberation observed at the i^{th} hydrophone is

$$V_i(t) = \sum_{\ell} \frac{a_\ell}{t_\ell^2} s_1(t-t_\ell) \cos\left[w_0(t-t_\ell) + \phi(t-t_\ell)\right] \tag{61}$$

t_ℓ is the travel time of sound from the origin of coordinates (nominal center of the source) to the ℓ^{th} scatterer and back to the i^{th} hydrophone. a_ℓ measures the amplitude of the signal reflected by the ℓ^{th} scatterer. It includes effects of the transmitter beam pattern as well as those of scatterer cross-section.[2] Similarly, the reverberation at the j^{th} hydrophone is

$$V_j(t) = \sum_{\ell} \frac{a_\ell}{T_\ell^2} s_1(t-T_\ell) \cos\left[w_0(t-T_\ell) + \phi(t-T_\ell)\right] \tag{62}$$

T_ℓ is the sound travel time from the origin to the j^{th} hydrophone via the ℓ^{th} scatterer. Hence the desired cross-correlation assumes the form

---

[1]The effect of strong target Doppler shifts has already been discussed in Section III.

[2]The general nomenclature is that of Report No. 27.

$$R_{ij}(\tau) = E\left\{\sum_{\ell}\sum_{m} \frac{a_\ell a_m}{t_\ell^2 T_m^2} s_1(t-t_\ell)\, s_1(t-T_m+\tau) \times \right.$$
$$\left. \times \cos[w_0(t-t_\ell) + \phi(t-t_\ell)] \cos[w_0(t-T_m+\tau) + \phi(t-T_m+\tau)]\right\} \quad (63)^1$$

Expressing the cos( ) cos( ) product in terms of sum and difference frequencies, one can invoke the Riemann-Lebesgue lemma to eliminate all but the difference terms of form $\ell = m$. Hence

$$R_{ij}(\tau) = \frac{1}{2} E\left\{\sum_{\ell} \frac{a_\ell^2}{t_\ell^2 T_\ell^2} s_1(t-t_\ell)s_1(t-T_\ell+\tau)\cos[w_0(t_\ell-T_\ell+\tau)+\phi(t-T_\ell+\tau)-\phi(t-t_\ell)]\right\} \quad (64)$$

The next step is to express T_ℓ in terms of t_ℓ. Consider the spherical coordinate system shown in Figure 2. By arbitrary convention

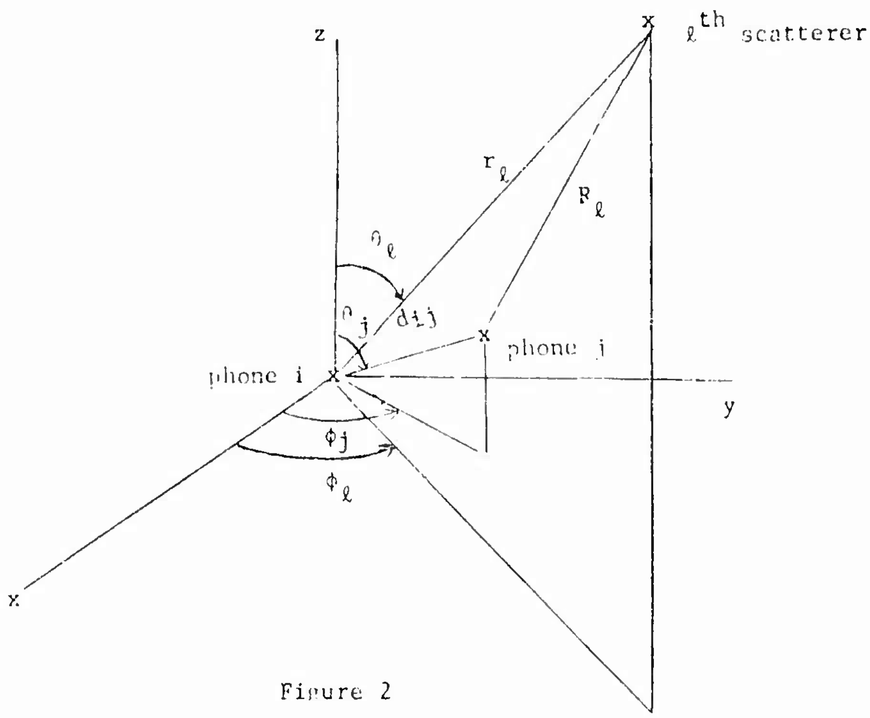


Figure 2

---

[1] The symbol E{ } denotes the expectation of the bracketed quantity.

the origin is placed at the i^{th} hydrophone. d_{ij} is the distance between phones i and j. A simple trigonometric computation now yields the distance R_j of the ℓ^{th} scatterer from phone j in terms of r_ℓ, the distance of the ℓ^{th} scatterer from phone i.

$$R_\ell = r_i \sqrt{1 - \frac{2d_{ij}}{r_\ell}\left[\sin\theta_\ell \sin\theta_j \cos(\phi_\ell - \phi_j) + \cos\theta_\ell \cos\theta_j\right] + \frac{d_{ij}^2}{r_\ell^2}} \tag{64}$$

In practice $\frac{d_{ij}}{r_\ell} << 1$ for all scatterers sufficiently close to the target to receive some illumination simultaneously with the target. Using this approximation and dividing both sides of Equation (64) by the velocity of sound one obtains

$$T_\ell = t_\ell - \frac{d_{ij}}{c}\left[\sin\theta_\ell \sin\theta_j \cos(\phi_\ell - \phi_j) + \cos\theta_\ell \cos\theta_j\right] \tag{65}$$

For greater ease in subsequent manipulation we introduce the notation

$$\alpha = \sin\theta_\ell \sin\theta_j \cos(\phi_\ell - \phi_j) + \cos\theta_\ell \cos\theta_j \tag{66}$$

Then

$$R_{ij}(\tau) = \frac{1}{2} E\left\{\sum_\ell \frac{a_\ell^2}{t_\ell^2 (t_\ell - \alpha \frac{d_{ij}}{c})^2} s_1(t-t_\ell) s_1(t-t_\ell+\alpha\frac{d_{ij}}{c}+\tau) \times\right.$$

$$\left. \cdot \cos\left[w_0(\alpha \frac{d_{ij}}{c} + \tau) + \phi(t-t_\ell+\alpha\frac{d_{ij}}{c}+\tau) - \phi(t-t_\ell)\right]\right\}, \tag{67}$$

The joint probability density of t_ℓ, ϕ_ℓ and θ_ℓ has been calculated

in Report No. 27. For volume reverberation from scatterers independently and uniformly distributed over a large volume V the result is

$$p(t_\ell, \theta_\ell, \phi_\ell) = \frac{c^3}{8V} t_\ell^{\,2} \sin \theta_\ell \tag{68}$$

The coefficient $a_\ell^{\,2}$ is now decomposed into its two primary components

$$a_\ell^{\,2} = \frac{b_\ell^{\,2}}{\sin\theta_\ell} g(\theta_\ell-\theta_0, \phi_\ell-\phi_0) \tag{69}$$

$b_\ell^{\,2}$ is proportional to the scattering cross-section while $g(\theta, \phi)$ is the transmitter pattern,[1] centered at (θ_0, ϕ_0) .

With the introduction of Equations (68) and (69) and the change of variable

$$x \equiv t - t_\ell + \alpha\frac{d_{ij}}{c} + \frac{T}{2} \tag{70}$$

Equation (67) becomes

$$R_{ij}(\tau) \equiv \frac{c^3}{16V} \overline{\sum_\ell \frac{b_\ell^{\,2}}{t_0^{\,2}}} \int_0^{2\pi} d\phi_\ell \int_0^{\pi} d\theta_\ell \, g(\theta_\ell-\theta_0, \phi_\ell-\phi_0) \int_{-\infty}^{\infty} dx s_1(x- \frac{T}{2} - \alpha\frac{d_{ij}}{c}) s_1(x+\frac{T}{2})$$

$$\times \cos\left[w_0(\alpha\frac{d_{ij}}{c} + \tau) + \phi(x + \frac{T}{2}) - \phi(x - \frac{T}{2} - \alpha\frac{d_{ij}}{c})\right] \tag{71}$$[2]

---

[1]The use of a frequency independent pattern function implies that the signal is sufficiently narrow-band to have its directional properties described by a single frequency pattern function.

[2]The bar indicates an averaging operation over the number of illuminated scatterers.

The slowly varying amplitude factor $b_i^2/\left(t_i - \alpha\frac{d_{ij}}{c}\right)^2$ has here been replaced with b_i^2/t_0^2, where t_0 [illegible] travel time from the target to the origin of coordinates. For targets remote compared with the radial distance covered by one pulse this should be an excellent approximation.

Two further approximations will now be made, both based on the observation that [illegible] is likely to be quite small. It is a simple matter to demonstrate that [illegible]. Hence, if the maximum distance between any pair of hydrophones in the array is 20 ft., $\left|\alpha\frac{d_{ij}}{c}\right| \leq 0.004$ sec. The typical pulse envelope function $s_1(t)$ does not change significantly over an interval of 4 milliseconds. Hence,

1) $s_1\left(x - \frac{\tau}{2} - \alpha\frac{d_{ij}}{c}\right)$ is approximated by $s_1(x - \frac{\tau}{2})$ in Equation (71).

$\phi(t)$ is also a relatively slowly varying function. Using a Taylor series one can write

$$\phi\left(x - \frac{\tau}{2} - \alpha\frac{d_{ij}}{c}\right) \approx \phi(x - \frac{\tau}{2}) + \phi'(x - \frac{\tau}{2})\,\alpha\frac{d_{ij}}{c} \tag{72}$$

$\phi'(t)$ is the instantaneous frequency modulation in radians/sec.

$\phi'(t)\,\alpha\frac{d_{ij}}{c}$ is the phase shift of the modulation frequency between hydrophones i and j [illegible].

2) $\phi\left(x - \frac{\tau}{2} - \alpha\frac{d_{ij}}{c}\right)$ is approximated by $\phi(x - \frac{\tau}{2})$ on the assumption that the phase shift at the maximum frequency deviation (from w_0) is small between any pair of hydrophones.

With a maximum deviation [illegible] bandwidth, if both positive

and negative deviations are allowed) and a maximum spacing between hydrophones of 20 ft., one finds that

$$\max \left[\phi'(x - \frac{\tau}{2})\, \alpha \frac{d_{ij}}{c} \right] \approx 0.4\pi \tag{73}$$

Approximation 2) has certainly become questionable for these parameters. However, in any practical array a very small percentage of the total hydrophone pairs have spacings close to the maximum array dimension. Furthermore, one tends to use arrays in such a manner that few, if any, hydrophone pairs assume an endfire alignment relative to the target. Hence, α will generally be well below unity and $\alpha \frac{d_{ij}}{c}$ in a 20 ft. array would be substantially below 0.004 sec for all, or almost all, pairs. Hence, approximation 2) appears not unreasonable for arrays of this general size and bandwidths up to the order of 100 cps .[1]

With approximations 1) and 2) Equation (71) becomes

$$R_{ij}(\tau) \cong \frac{c^3}{16V} \sum_{\ell} \overline{\frac{b_\ell^2}{t_0^2}} \int_0^{2\pi} d\phi_\ell \int_0^{\pi} d\theta_\ell\, g(\theta_\ell - \theta_0, \phi_\ell - \phi_0) \int_{-\infty}^{\infty} dx\, s_1(x - \frac{\tau}{2})\, s_1(x + \frac{\tau}{2}) \times$$

$$\cos\left[w_0 \left(\alpha \frac{d_{ij}}{c} + \tau \right) + \phi(x + \frac{\tau}{2}) - \phi(x - \frac{\tau}{2}) \right] \tag{74}$$

Assuming, for the sake of computational simplicity, that $s_1(\ )$ and $\phi(\ )$ are both even, i.e.,

---

[1] By introducing an assumption of narrow beam patterns one can, in fact, considerably weaken the inequality $\left| \phi'(x - \frac{\tau}{2})\, \alpha \frac{d_{ij}}{c} \right| << 1$ implied by 2) . The appropriate weaker inequality is worked out in Appendix B.

$$s_1(x) = s_1(-x) \qquad \text{pulse envelope symmetrical} \tag{75}$$

and

$$\phi(x) = \phi(-x) \qquad \text{even phase modulation or odd frequency modulation, such as linear FM} \tag{76}$$

one can readily reduce Equation (74) to the form

$$R_{ij}(\tau) = \frac{C^3}{16V} \sum_{\ell} \frac{\overline{b_\ell^{\,2}}}{t_0^{\,2}} \int_0^{2\pi} d\phi_\ell \int_0^{\pi} d\theta_\ell \, g(\theta_\ell - \theta_0, \phi_\ell - \phi_0) \cos w_0\left(\alpha \frac{d_{ij}}{c} + \tau\right) \times$$

$$\int_{-\infty}^{\infty} dx \; s_1\left(x - \frac{\tau}{2}\right) s_1\left(x + \frac{\tau}{2}\right) \cos\left[\phi\left(x + \frac{\tau}{2}\right) - \phi\left(x - \frac{\tau}{2}\right)\right]$$

$$= \frac{C^3}{16V}\, p(\tau) \sum_{\ell} \frac{\overline{b_\ell^{\,2}}}{t_0^{\,2}} \int_0^{2\pi} d\phi_\ell \int_0^{\pi} d\theta_\ell \, g(\theta_\ell - \theta_0, \phi_\ell - \phi_0) \cos w_0\left(\alpha \frac{d_{ij}}{c} + \tau\right) \tag{77}$$

where

$$p(\tau) \equiv \int_{-\infty}^{\infty} dx \; s_1\left(x - \frac{\tau}{2}\right) s_1\left(x + \frac{\tau}{2}\right) \cos\left[\phi\left(x + \frac{\tau}{2}\right) - \phi\left(x - \frac{\tau}{2}\right)\right] \tag{78}$$

The double integral in Equation (77) depends only on transmitter and receiver geometry, while $p(\tau)$ depends only on the waveshape of the transmitted signal. The complete separation of these two effects is the direct consequence of approximations 1) and 2).

The double integral can be further simplified if one considers symmetrical pattern functions narrowly concentrated near (θ_0, ϕ_0). In that case one can extend the ϕ_ℓ and θ_ℓ limits to $(-\infty, \infty)$ and represent α [Equation (66)] by the first three terms of a Taylor series

$$\alpha \cong \alpha_{ij} + \beta_{ij}(\theta_\ell - \theta_0) + \gamma_{ij}(\phi_\ell - \phi_0) \tag{79}$$

where

$$\alpha_{ij} = \sin\theta_0 \sin\theta_j \cos(\phi_0 - \phi_j) + \cos\theta_0 \cos\theta_j \tag{80}$$

$$\beta_{ij} = \cos\theta_0 \sin\theta_j \cos(\phi_0 - \phi_j) - \sin\theta_0 \cos\theta_j \tag{81}$$

$$\gamma_{ij} = \sin\theta_0 \sin\theta_j \sin(\phi_j - \phi_0) \tag{82}$$

With the changes of variables $\theta_\ell - \theta_0 = u$, $\phi_\ell - \phi_0 = v$ the double integral becomes

$$\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv\ g(u, v) \cos w_0 \left[\tau + \frac{d_{ij}}{c}\left(\alpha_{ij} + \beta_{ij}\, u + \gamma_{ij}\, v\right)\right] \tag{83}$$

According to Equation (20) we are ultimately interested in the value of the autocorrelation function not at τ, but at $t - \lambda - \tau_i + \tau_j$. However, with the array steered on target we obtain from Equations (65) and (66)

$$\tau_i - \tau_j = \frac{d_{ij}}{c}\alpha_{ij} \tag{84}$$

Hence, from Equations (77), (83) and (84), using the postulated symmetry of the pattern function about (θ_0, ϕ_0)

$$P_{ij}(\tau - \tau_i + \tau_j) = \frac{c^3}{16V} p(\tau) \cos w_0\tau \overline{\sum_\ell \frac{b_\ell^2}{t_0^2}} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv\ g(u, v) \cos w_0 \frac{d_{ij}}{c}(\beta_{ij} u + \gamma_{ij} v)$$

$$= \frac{c^3}{16V}\overline{\left(\sum_\ell \frac{b_\ell^2}{t_0^2}\right)} G\left[\left(\frac{d_{ij}}{c}\beta_{ij} w_0\right), \left(\frac{d_{ij}}{c}\gamma_{ij} w_0\right)\right] p(\tau) \cos w_0\tau \tag{85}$$

where

$$G(w, z) = \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv\ g(u, v) \cos (wu + zv)$$

$$= \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv\ g(u, v)\ e^{-j(wu + zv)} \tag{86}$$

Thus $G(w, z)$ is the Fourier transform of the pattern function $g(u, v)$.

To normalize the crosscorrelation function we need only recognize from Equation (85) that

$$R_{ii}(0) = R_{jj}(0) = \frac{c^3}{16V} \overline{\left(\sum_{\ell} \frac{b_\ell^2}{t_0^2} \right)} G(0, 0)\ p(0) \tag{87}$$

Hence,

$$\rho_{ij}(\tau-\tau_i+\tau_j) = \frac{G\left[\left(\frac{d_{ij}}{c}\beta_{ij}w_0\right), \left(\frac{d_{ij}}{c}\gamma_{ij}w_0\right)\right]}{G(0, 0)} \frac{p(\tau)}{p(0)} \cos w_0 \tau \tag{88}$$

Equation (88) must now be substituted into Equation (60). Designating the general term of Equation (60) by A_n as before and using an obvious generalization of the steps leading to Equation (38) one finds

$$\frac{A_n}{K_n C_{1n}} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} \left\{ \frac{G\left[\left(\frac{d_{ij}}{c}\beta_{ij}w_0\right), \left(\frac{d_{ij}}{c}\gamma_{ij}w_0\right)\right]}{G(0, 0)} \right\}^n \int_{-\infty}^{\infty} dt\ s_1(t) \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda) \cos\left[\phi(t)-\phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n}{\sum_{i=1}^{M} \sum_{j=1}^{M} \frac{G\left[\left(\frac{d_{ij}}{c}\beta_{ij}w_0\right), \left(\frac{d_{ij}}{c}\gamma_{ij}w_0\right)\right]}{G(0, 0)} \int_{-\infty}^{\infty} dt\ s_1(t) \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda) \cos\left[\phi(t)-\phi(\lambda)\right] \frac{p(t-\lambda)}{p(0)}} \tag{89}$$

$g(u, v)$ describes the spatial distribution of radiated power and is therefore non-negative. It follows from Equation (86) that

$$G(u, v) \leq G(0, 0) \quad \text{for all} \quad u, v \tag{90}$$

Furthermore, it is shown in Appendix C that for narrow beam patterns

$$G(u, v) \geq 0 \tag{91}$$

The ratio of the double sums Equation (89) is therefore no larger than unity and one obtains

$$A_n \leq K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dt\ s_1(t) \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda)\ \cos\left[\phi(t)-\phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n}{\displaystyle\int_{-\infty}^{\infty} dt\ s_1(t) \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda)\ \cos\left[\phi(t)-\phi(\lambda)\right] \frac{p(t-\lambda)}{p(0)}} \tag{92}$$

Using Equations (75) and (76) it is a simple matter to show further that

$$A_n \leq K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dt\ s_1(t)\ e^{j\phi(t)} \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda)\ e^{-j\phi(\lambda)} \left[\frac{p(t-\lambda)}{p(0)}\right]^n}{\displaystyle\int_{-\infty}^{\infty} dt\ s_1(t)\ e^{j\phi(t)} \int_{-\infty}^{\infty} d\lambda\ s_1(\lambda)\ e^{-j\phi(\lambda)} \frac{p(t-\lambda)}{p(0)}} \tag{93}$$

The fact that both double integrals are in form of convolutions suggests the use of Fourier transforms. Define

$$S_{\ell f}(w) = \int_{-\infty}^{\infty} s_1(t)\ e^{j\phi(t)}\ e^{-jwt}\ dt \tag{94}$$

and

$$P(w) = \int_{-\infty}^{\infty} p(t)\ e^{-jwt}\ dt \tag{95}$$

Thus $S_{\ell f}(w)$ is the Fourier transform of the low frequency signal, i.e., the signal after a downward shift by w_o. In terms of Equations (94) and (95) A_n can now be written as follows

$$A_n \leq K_n C_{1n} |p(0)|^{-(n-1)} \frac{\int_{-\infty}^{\infty} dw\ |S_{\ell f}(w)|^2 \{P(w) * P(w) * \ldots * P(w)\}}{\int_{-\infty}^{\infty} dw\ |S_{\ell f}(w)|^2\ P(w)} \tag{96}$$

Here $A * B$ denotes the convolution of A with B.

Consider next the relation between $P(w)$ and $S_{\ell f}(w)$. Using the postulated symmetry of s_1 and ϕ one obtains from Equation (78)

$$p(\tau) = \int_{-\infty}^{\infty} dx\ s_1(x - \tfrac{\tau}{2})\ s_1(x + \tfrac{\tau}{2})\ e^{j\phi(x + \frac{\tau}{2})}\ e^{-j\phi(x - \frac{\tau}{2})} \tag{97}$$

The change of variable $x + \frac{\tau}{2} = y$ leads to

$$p(\tau) = \int_{-\infty}^{\infty} dy\ s_1(y)\ e^{j\phi(y)}\ s_1(y - \tau)\ e^{-j\phi(y-\tau)} \tag{98}$$

which is a convolution of $s_1(y)\ e^{j\phi(y)}$ with $s_1(y)\ e^{-j\phi(y)}$. It follows that

$$P(w) = |S_{\ell f}(w)|^2 \tag{99}$$

Thus, $P(w)$ is real. Now substituting Equation (99) into Equation (96) and using Parseval's theorem

$$A_n \leq K_n C_{1n} [p(0)]^{-(n-1)} \frac{\int_{-\infty}^{\infty} dw \; P(w) \; \{P(w) * P(w) * \ldots * P(w)\}}{\int_{-\infty}^{\infty} dw \; P(w) \cdot P(w)}$$

$$= K_n C_{1n} \frac{\int_{-\infty}^{\infty} d\tau \left[\frac{p(\tau)}{p(o)}\right]^{n+1}}{\int_{-\infty}^{\infty} d\tau \left[\frac{p(\tau)}{p(0)}\right]^{2}} \tag{100}$$

Since n assumes only odd values, the integrands of both numerator and denominator are non-negative. Furthermore, from Equation (99)

$$|p(\tau)| = \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} dw |S_{\ell f}(w)|^2 \; e^{jw\tau} \right| \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} dw |S_{\ell f}(w)|^2 = p(0) \tag{101}$$

Hence, the ratio of integrals in Equation (100) has an upper bound of unity and

$$A_n \leq K_n C_{1n} \tag{102}$$

But this is identical with Equation (39) so that one obtains immediately from Equation (44)

$$R \geq 0.89 \tag{103}$$

Thus, at least in the absence of Doppler shifts, clipping losses in a reverberation limited environment are quite small for a very general class of signals and arrays.

V. Range Estimation

An estimate of target range can be obtained by regarding the correlator output as a function of replica delay (τ) and establishing the "location" of this function on the τ axis. Details of the required instrumentation depend on the precise definition of the term "location", but if the signal correlation function is sufficiently concentrated in τ to permit useful range estimates one would expect any reasonable measure of "location" to lead to comparable results. One such measure is obtained by the arrangement shown in Figure 3. $x(t)$ is the output of the beamformer as in Figure 1.

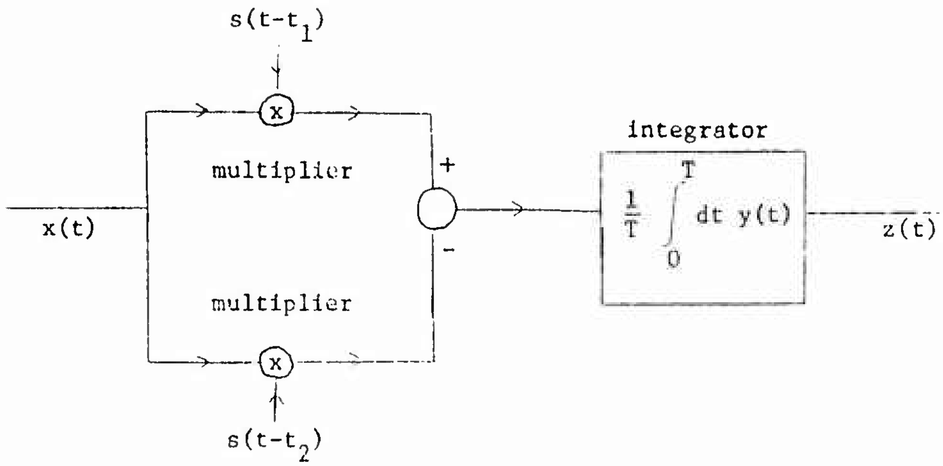


Figure 3

It is cross-correlated with two replicas of the signal, delayed by t_1 and t_2 seconds respectively. The resulting short time correlation functions are subtracted to yield the final output $z(t)$. If the target delay (t_0) is given by

$$t_0 = \frac{t_1 + t_2}{2} \qquad (104)$$

the expected value of z is zero. Deviations of $\overline{z}$ from zero indicate

values of target delay other than that given in Equation (104). The measurement error of such an instrumentation has been discussed in Report No. 29 (Section I). If the true target delay is t_0 then the rms measurement error is

$$\sigma_{t_0} = \frac{\sqrt{\overline{z^2}\Big|_{t_0 = \frac{t_1+t_2}{2}}}}{\frac{\partial \bar{z}}{\partial t_0}\Big|_{t_0 = \frac{t_1+t_2}{2}}} \tag{105}$$ [1]

This figure of merit must now be calculated for instrumentations with and without clipping. In the absence of clipping the equivalent of Equation (2) is

$$x(t) = M\, s(t-t_o) + \sum_{i=1}^{M} n_i(t-\tau_i) \tag{106}$$ [2]

Hence,

$$y_1(t) = M\, s(t-t_0)\, s(t-t_1) + \sum_{i=1}^{M} n_i(t-\tau_i)\, s(t-t_1) \tag{107}$$

---

[1]For sufficiently high signal-to-noise ratio to make meaningful range measurement possible.

[2]The target delay t_0 was omitted in Equation (1). Since range was assumed to be known in the detection study all delays could be measured relative to t_0 .

and

$$y_2(t) = M\, s(t-t_0)\, s(t-t_2) + \sum_{i=1}^{M} n_i(t-\tau_i)\, s(t-t_2) \tag{108}$$

Since the average value of the noise is zero

$$\overline{z} = \frac{M}{T} \int_0^T dt\, s(t-t_0) \left[s(t-t_1) - s(t-t_2) \right] \tag{109}$$

Similarly in the presence of clipping one obtains from a computation parallel to Equations (2) - (10) [low input signal-to-noise ratio]

$$\overline{z} = \sqrt{\frac{2}{\pi}}\, \frac{M}{T\sqrt{N}} \int_0^T dt\, s(t-t_0) \left[s(t-t_1) - s(t-t_2) \right] \tag{110}$$

Hence, by a trivial computation

$$\frac{\left.\dfrac{\partial \overline{z}}{\partial t_0}\right|_{\text{clipped}}}{\left.\dfrac{\partial \overline{z}}{\partial t_0}\right|_{\text{unclipped}}} = \sqrt{\frac{2}{\pi N}} \tag{111}$$

In the absence of clipping the mean square value of z is

$$\overline{z^2} = \left[\frac{1}{T} \int_0^T dt\, s(t-t_1) \sum_{i=1}^{M} n_i(t-\tau_i) - \frac{1}{T} \int_0^T dt\, s(t-t_2) \sum_{i=1}^{M} n_i(t-\tau_i) \right]^2$$

$$= \frac{1}{T^2} \int_0^T dt \int_0^T d\lambda \left[s(t-t_1)s(\lambda-t_1) - 2s(t-t_1)s(\lambda-t_2) + s(t-t_2)s(\lambda-t_2) \right] \times$$

$$\times\, N \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}(t-\lambda-\tau_i+\tau_j) \tag{112}$$[1]

---

[1]The contribution of the signal fluctuations to $\overline{z^2}$ has been ignored under the assumption of low input signal-to-noise ratio.

Similarly in the presence of clipping [see Equations (11) - (13)]

$$\overline{z^2} = \frac{1}{T^2}\int_0^T dt \int_0^T d\lambda\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right] \times$$

$$\times \frac{2}{\pi}\sum_{i=1}^{M}\sum_{j=1}^{M}\sin^{-1}\left[\rho_{ij}(t-\lambda-\tau_i+\tau_j)\right] \tag{113}$$

From Equations (105) and (111) - (113) the clipping loss is

$$R = \frac{\sigma_{t_0}\Big|_{\text{clipped}}}{\sigma_{t_0}\Big|_{\text{unclipped}}}$$

$$= \left\{\frac{\int_0^T dt \int_0^T d\lambda\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]}{\int_0^T dt \int_0^T d\lambda\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]} \cdot \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\rho_{ij}(t-\lambda-\tau_i+\tau_j)}{\sum_{i=1}^{M}\sum_{j=1}^{M}\sin^{-1}\left[\rho_{ij}(t-\lambda-\tau_i+\tau_j)\right]}\right\}^{\frac{1}{2}} \tag{114}$$

The similarity with Equation (20) is obvious. One can clearly carry through many of the general arguments of Section III and obtain similar results. Here we shall concern ourselves only with the case of reverberation in the absence of Doppler shift [treated in Section IV, Equations (61) - (93)]. Once again we work with the general narrowband signal

$$s(t) = s_1(t)\left[\cos w_0 t + \phi(t)\right] \tag{115}$$

With the same restrictions on array dimensions as in Section IV [array diameter small compared with wavelength of maximum frequency deviation from w_0], one obtains from Equations (60) and (88)

$$A_n = K_n \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\left\{\frac{G\left[\left(\frac{d_{ij}}{c}\beta_{ij}w_0\right),\left(-\frac{d_{ij}}{c}\gamma_{ij}w_0\right)\right]}{G(0,\ 0)}\right\}^n \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda}{\sum_{i=1}^{M}\sum_{j=1}^{M}\left\{\frac{G\left[\left(\frac{d_{ij}}{c}\beta_{ij}w_0\right),\left(\frac{d_{ij}}{c}\gamma_{ij}w_0\right)\right]}{G(0,\ 0)}\right\} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda}$$

$$\frac{\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]\left[\frac{p(t-\lambda)}{p(0)}\right]^n \cos^n w_0(t-\lambda)}{\left[s(t-t_1)s(\lambda-t_2)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]\frac{p(t-\lambda)}{p(0)} \cos w_0(t-\lambda)} \tag{116}$$

where K_n is defined by Equation (25), $p(t)$ by Equation (78), and A_n is the n^{th} term of the expansion of $1/R^2$. The limits of integration have been extended to $(-\infty, \infty)$ on the assumption that the integration time $(0, T)$ at least covers the duration of the two replicas. As in Section IV, the ratio of the double sums has an upper bound of unity. Hence,

$$A_n \le K_n \frac{\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]}{\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda\left[s(t-t_1)s(\lambda-t_1)-2s(t-t_1)s(\lambda-t_2)+s(t-t_2)s(\lambda-t_2)\right]}$$

$$\frac{\left[\frac{p(t-\lambda)}{p(0)}\right]^n \cos^n w_0(t-\lambda)}{\frac{p(t-\lambda)}{p(0)} \cos w_0(t-\lambda)} \tag{117}$$

Only the term in $s(t-t_1)\ s(t-t_2)$ differs from the forms treated previously. One can express the constraint of Equation (104) by

$$t_2 = t_0 + \Delta \tag{118}$$

$$t_1 = t_0 - \Delta \tag{119}$$

Substituting Equations (115), (118) and (119) into (117), one obtains after some algebraic manipulation (invoking the Riemann-Lebesgue lemma)

$$A_n \leq K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \{s_1(t)s_1(\lambda)\cos\left[\phi(t)-\phi(\lambda)\right]-s_1(t+\Delta)s_1(\lambda-\Delta)\cos\left[2w_0\Delta+\phi(t+\Delta)-\phi(\lambda-\Delta)\right]\}\left|\frac{p(t-\lambda)}{p(0)}\right|^n}{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \{s_1(t)s_1(\lambda)\cos\left[\phi(t)-\phi(\lambda)\right]-s_1(t+\Delta)s_1(\lambda-\Delta)\cos\left[2w_0\Delta+\phi(t+\Delta)-\phi(\lambda-\Delta)\right]\}\frac{p(t-\lambda)}{p(0)}} \tag{120}$$

The signal correlation function is monotone only over intervals of the order of half a carrier cycle. Hence, the separation (2Δ) between the two replica delays cannot exceed π/w_0 if the output from the device of Figure 3 is to have an unambiguous interpretation.[1] For time increments of the order of π/w_0 the relatively slowly varying functions $s_1(t)$ and $\phi(t)$ do not change significantly. Hence,

---

[1]The practically more interesting situation in which only the envelope of the correlation function is used in ranging is discussed on p. 42.

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t+\Delta) s_1(\lambda-\Delta) \cos\left[2w_0\Delta + \phi(t+\Delta) - \phi(\lambda-\Delta)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$\cong \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \cos\left[2w_0\Delta + \phi(t) - \phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$= \cos 2w_0\Delta \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \cos\left[\phi(t) - \phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$- \sin 2w_0\Delta \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \sin\left[\phi(t) - \phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \tag{121}$$

The change of variable

$$t - \lambda = x$$
$$t + \lambda = y \tag{122}$$

yields

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \sin\left[\phi(t) - \phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$= \tfrac{1}{2} \int_{-\infty}^{\infty} dx \left[\frac{p(x)}{p(0)}\right]^n \int_{-\infty}^{\infty} dy s_1\left(\frac{x+y}{2}\right) s_1\left(\frac{y-x}{2}\right) \sin\left[\phi\left(\frac{x+y}{2}\right) - \phi\left(\frac{y-x}{2}\right)\right] \tag{123}$$

s_1 and ϕ are even functions by assumption. Therefore $\phi\left(\frac{y+x}{2}\right) - \phi\left(\frac{y-x}{2}\right)$ is odd in y. It follows that the integrand of Equation (123) is odd in y so that the value of the integral is zero. Using this result in Equation (121) and substituting in Equation (120) one obtains finally

$$A_n \le K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \cos\left[\phi(t) - \phi(\lambda)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n}{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t) s_1(\lambda) \cos\left[\phi(t) - \phi(\lambda)\right] \frac{p(t-\lambda)}{p(0)}} \tag{124}$$

But this is identical with Equation (92) so that Equation (103) remains true:

$$R \geq 0.89 \tag{125}$$

The discussion just concluded is unrealistic in one respect. The postulated instrumentation uses the carrier frequency in ranging, thus obtaining in effect extreme range accuracy (to a small fraction of a carrier wavelength) at the expense of ambiguity over multiples of the carrier wavelength. In practice one cannot tolerate such ambiguity. One would therefore almost certainly ignore carrier frequency effects and seek to locate the envelope of the signal correlation function on the τ axis. The formal analysis of an appropriate instrumentation is fairly cumbersome, but much of the desired insight into the question of clipping loss can be obtained from the following line of reasoning.

A typical signal autocorrelation function is sketched in Figure 4.

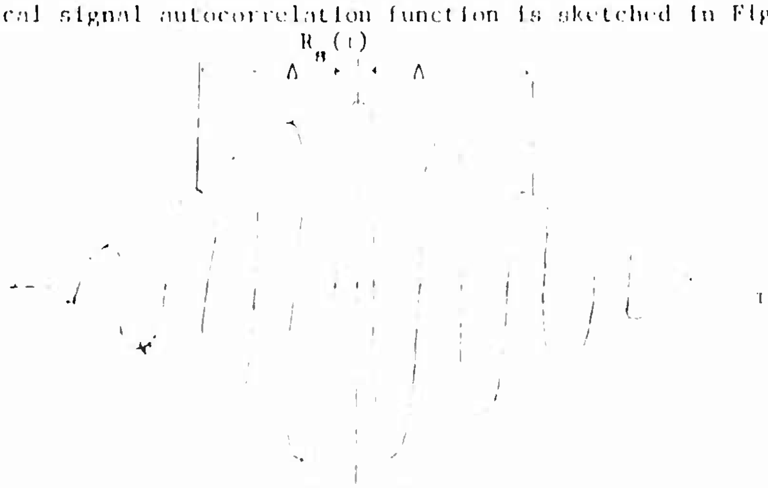


Figure 4

an instrumentation of the type of Figure 3 designed to track the envelope would have to compare the amplitude of the quasi-sinusoidal oscillations at a distance Δ from the origin. This clearly leads to lower sensitivities $\frac{\partial z}{\partial t_0}$ than in the previous computation, where Δ was of the order of a quarter wavelength of the carrier frequency. On the other hand, it is clear

from Equations such as (109) and (110) that the ratio of sensitivities for the clipped and unclipped instrumentations [Equation (111)] is unaffected by this change. Furthermore, the output fluctuation $\overline{z^2}$ for the envelope observation is simply the output fluctuation of Figure 3 with the delay adjusted for operation at the peak of the appropriate carrier cycle (as suggested in Figure 4). Thus, Equation (114) and hence, Equation (120) is still indicative of the clipping loss if Δ assumes the appropriate value.[1] One can now no longer make the approximation in the second line of Equation (121) and must write instead

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \; s_1(t+\Delta)\, s_1(\lambda-\Delta)\, \cos\left[2w_0\Delta + \phi(t+\Delta) - \phi(\lambda-\Delta)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$= \cos 2w_0\Delta \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \; s_1(t+\Delta)\, s_1(\lambda-\Delta)\, \cos\left[\phi(t+\Delta) - \phi(\lambda-\Delta)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n$$

$$- \sin 2w_0\Delta \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \; s_1(t+\Delta)\, s_1(\lambda-\Delta)\, \sin\left[\phi(t+\Delta) - \phi(\lambda-\Delta)\right] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \tag{126}$$

The change of variable (122) applied to the last term of Equation (126) still leads to an odd y function so that this integral vanishes. The equivalent of Equation (124) is therefore

---

[1] The zeros of $R_s(\tau)$ occur at values of τ such that $w_0\tau = K\frac{\pi}{2}$, K odd. However, the peaks of $R_s(\tau)$ do not necessarily occur midway between the zeros, a fact which causes an apparent difficulty in the precise choice of Δ. Fortunately, it turns out that changes of Δ by a small fraction of a carrier cycle do not affect the clipping loss computation.

$$\Lambda_n \le K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \left|\frac{p(t-\lambda)}{p(0)}\right|^n \{s_1(t)s_1(\lambda)\cos[\phi(t)-\phi(\lambda)] - \cos 2w_0\Delta s_1(t+\Delta)s_1(\lambda-\Delta)\cos[\phi(t+\Delta)-\phi(\lambda-\Delta)]\}}{\displaystyle\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda \frac{p(t-\lambda)}{p(0)} \{s_1(t)s_1(\lambda)\cos[\phi(t)-\phi(\lambda)] - \cos 2w_0\Delta s_1(t+\Delta)s_1(\lambda-\Delta)\cos[\phi(t+\Delta)-\phi(\lambda-\Delta)]\}} \tag{127}$$

The first terms in numerator and denominator are identical with the integrals in Equation (92). The second terms can be reduced to an analogous form by the transformation

$$\begin{aligned} x &= t + \Delta \\ y &= t - \Delta \end{aligned} \tag{128}$$

Following the same sequence of steps as in Equations (92) - (100) one now obtains

$$\Lambda_n \le K_n C_{1n} \frac{\displaystyle\int_{-\infty}^{\infty} d\tau \left\{\left|\frac{p(\tau)}{p(0)}\right|^{n+1} - \cos 2w_0\Delta \left|\frac{p(\tau)}{p(0)}\right|^n \frac{p(\tau-2\Delta)}{p(0)}\right\}}{\displaystyle\int_{-\infty}^{\infty} d\tau \left\{\left|\frac{p(\tau)}{p(0)}\right|^2 - \cos 2w_0\Delta \left|\frac{p(\tau)}{p(0)}\right| \frac{p(\tau-2\Delta)}{p(0)}\right\}} \tag{129}$$

Because of the negative terms it is difficult to draw conclusions of the same generality as in the case of detection [Equations (100) - (102)]. The practical picture becomes clear, however, when one recalls that $p(\tau)$ [Equation (97)] is in effect the autocorrelation function $R_s(\tau)$ of the signal. Furthermore, Δ must be chosen in a range of τ values where $R_s(\tau)$ decays rapidly. If the envelope of $R_s(\tau)$ is sufficiently

concentrated on the τ axis to permit meaningful range measurements $p(2\Delta)$ should be very small. Hence, the product $\left|\frac{p(\tau)}{p(0)}\right|^n \frac{p(\tau-2\Delta)}{p(0)}$ should be small for all τ and the second terms in numerator and denominator of Equation (129) should have only a minor effect on the ratio of integrals. Thus, one expects Equation (102) to remain true with, at most, minor modifications.

As an example, consider a signal consisting of a Gaussian pulse with linear frequency modulation

$$s(t) = e^{-\frac{t^2}{\sigma_T^2}} \cos(w_0 t + \frac{K}{2} t^2) \tag{130}$$

A straightforward computation yields

$$\frac{p(\tau)}{p(0)} = e^{-\Omega^2 \tau^2} \tag{131}$$

where

$$\Omega = \sqrt{\frac{1}{2\sigma_T^2} + \frac{K^2 \sigma_T^2}{8}} \tag{132}$$

Thus, Ω is the effective bandwidth of the transmitted signal. Substituting Equation (131) into Equation (129) one obtains

$$A_n \leq K_n C_{1n} \sqrt{\frac{2}{n+1}} \; \frac{1 - \cos 2w_0\Delta \, e^{-\frac{n}{n+1} 4 \Omega^2 \Delta^2}}{1 - \cos 2w_0\Delta \, e^{-2\Omega^2\Delta^2}} \tag{133}$$

$p(\tau)/p(0)$ has maximum slope at $\tau = \pm \frac{1}{\sqrt{2}} \frac{1}{\Omega}$. Hence, this value of delay is chosen for Δ. Then

$$A_n \le K_n C_{1n} \sqrt{\frac{2}{n+1}} \; \frac{1 - \cos\left(\sqrt{2}\,\frac{w_0}{\Omega}\right) e^{-2\frac{n}{n+1}}}{1 - \cos\left(\sqrt{2}\,\frac{w_0}{\Omega}\right) e^{-1}} \tag{134}$$

The last fraction in Equation (134) has a maximum value when the cosine terms have arguments equal to multiples of 2π and when $n \to \infty$. Therefore

$$A_n \le K_n C_{1n} \sqrt{\frac{2}{n+1}} \; \frac{1 - e^{-2}}{1 - e^{-1}} = 1.37 \; K_n C_{1n} \sqrt{\frac{2}{n+1}} \tag{135}$$

n assumes only odd values. For $n = 3$

$$1.37 \; K_n C_{1n} \sqrt{\frac{2}{n+1}} = 0.966 \quad K_3 C_{13} \tag{136}$$

It follows that

$$A_n \le K_n C_{1n} \quad \text{for all} \quad n \ge 3 \tag{137}$$

Since $A_1 = 1$, 0.89 remains a lower bound on R. In this particular example the lower bound is definitely not reached, so that the clipping loss is actually even smaller.

VI. Doppler Estimation

Radial target velocity can be estimated by a procedure very similar to the one used in range estimation. Figure 5 shows a schematic Doppler estimator equivalent to the range estimator of Figure 3. The target delay is assumed to be known and the signal is

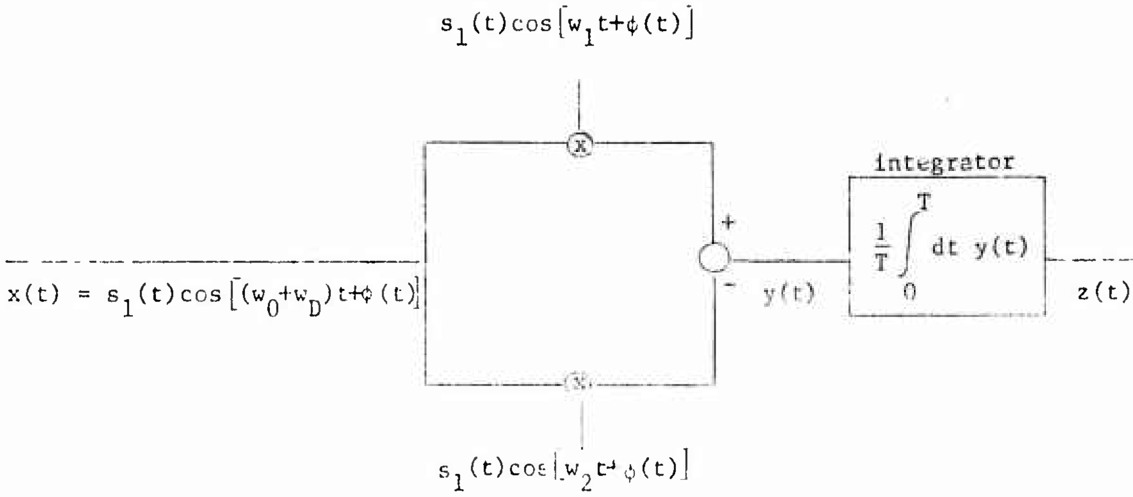


Figure 5

taken as sufficiently narrow-band so that Doppler shifts in the envelope and phase modulation may be ignored.

By analogy with Equation (105) the rms Doppler error is

$$\sigma_{w_D} = \frac{\sqrt{\overline{z^2}\,\Big|_{w_0+w_D = \frac{w_1+w_2}{2}}}}{\frac{\partial \bar{z}}{\partial w_0}\Big|_{w_0+w_D = \frac{w_1+w_2}{2}}} \tag{138}$$

A computation entirely parallel to Equations (106) - (111) verifies that

$$\frac{\frac{\partial \bar{z}}{\partial w_0}\Big|_{\text{clipped}}}{\frac{\partial \bar{z}}{\partial w_0}\Big|_{\text{unclipped}}} = \sqrt{\frac{2}{\pi N}} \tag{139}$$

Proceeding as in Section IV one obtains in place of Equation (117)

$$A_n \le K_n \frac{D_n + E_n - 2F_n}{D_1 + E_1 - 2F_1} \tag{140}$$

where

$$D_n = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t)s_1(\lambda)\cos[w_1 t+\phi(t)]\cos[w_1\lambda+\phi(\lambda)] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \cos^n w_0(t-\lambda) \tag{141}$$

$$E_n = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t)s_1(\lambda)\cos[w_2 t+\phi(t)]\cos[w_2\lambda+\phi(\lambda)] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \cos^n w_0(t-\lambda) \tag{142}$$

$$F_n = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t)s_1(\lambda)\cos[w_1 t+\phi(t)]\cos[w_2\lambda+\phi(\lambda)] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \cos^n w_0(t-\lambda) \tag{143}$$

Set

$$\begin{aligned} w_1 &= w_0+w_D - \Delta w \\ w_2 &= w_0+w_D + \Delta w \end{aligned} \tag{144}$$

Then a simple computation yields

$$D_n+E_n = \frac{C_{1n}}{2} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\lambda s_1(t)s_1(\lambda)\cos\Delta w(t-\lambda)\cos[w_D(t-\lambda)+\phi(t)-\phi(\lambda)] \left[\frac{p(t-\lambda)}{p(0)}\right]^n \tag{145}$$

With the change of variable

$$\begin{aligned} t - \lambda &= x \\ t + \lambda &= y \end{aligned} \tag{146}$$

one obtains (using the even symmetry of ϕ)

$$D_n+E_n = \frac{C_{1n}}{4} \int_{-\infty}^{\infty} dx\left[\frac{p(x)}{p(0)}\right]^n \cos\Delta wx\cos w_D x \int_{-\infty}^{\infty} dys_1\left(\frac{y+x}{2}\right)s_1\left(\frac{y-x}{2}\right)\cos\left[\phi\left(\frac{y+x}{2}\right)-\phi\left(\frac{y-x}{2}\right)\right]$$

$$= \frac{C_{1n}}{2} \int_{-\infty}^{\infty} dx\left[\frac{p(x)}{p(0)}\right]^{n+1} p(0)\cos\Delta wx\cos w_D x \tag{147}$$

A similar sequence of steps yields

$$F_n = \frac{C_{1n}}{8} \int_{-\infty}^{\infty} dx \left[\frac{p(x)}{p(0)}\right]^n \cos w_D x \int_{-\infty}^{\infty} dy\, s_1\left(\frac{y+x}{2}\right) s_1\left(\frac{y-x}{2}\right) \cos\left[\Delta wy - \phi\left(\frac{y+x}{2}\right) + \phi\left(\frac{y-x}{2}\right)\right] \tag{148}$$

$\frac{p(x)}{p(0)} \leq 1$. Hence $\left|\frac{p(x)}{p(0)}\right|^n$ decays more rapidly than $\frac{p(x)}{p(0)}$. It follows from Equation (147) that the ratio on the right side of Equation (140) can be made much larger than unity by choosing w_D sufficiently large. This, of course, is simply the phenomenon discussed in Section III: If the target Doppler shift is large there will indeed be substantial clipping loss because signal and noise spectra in the unclipped instrumentation are almost disjoint. We are here concerned with any additional loss which may occur in Doppler measurement. Hence, we choose $w_D = 0$ and attempt to set bounds on A_n for that case.

From a qualitative point of view one can make the following observations: Δw would be chosen so as to place w_1 and w_2 near the points of maximum slope of the signal spectrum. Roughly speaking this identifies Δw with the half-bandwidth of the signal spectrum. The effective duration of $\frac{p(x)}{p(0)}$ is given by the correlation time of the signal. Hence, the maximum value of Δwx in the effective range of integration is of the order of a radian for D_1+E_1 and less for D_n+E_n, $n > 1$. Therefore, the expression for D_n+E_n does not differ greatly from

$$\frac{C_{1n}}{2} \int_{-\infty}^{\infty} dx \left|\frac{p(x)}{p(0)}\right|^{n+1} p(0) \tag{149}$$

which is (except for a constant independent of n) the same as the numerator of Equation (100). In the expression for F_n, on the other

hand, the effective range of the y integration covers the signal duration and over that range Δwy could go through many complete periods (except in the absence, or virtual absence, of frequency modulation). From the Riemann-Lebesgue lemma one would then infer that $2F_n$ is small compared with D_n+E_n, except possibly in the absence of frequency modulation. One would therefore generally expect the bound on A_n given by Equation (140) not to differ drastically from Equation (137).

To give some quantitative support to this line of reasoning, consider once more the Gaussian pulse with linear frequency modulation

$$s(t) = e^{-\frac{t^2}{\sigma_T^2}} \cos(w_0 t + \frac{K}{2} t^2) \tag{150}$$

Straightforward computations yield

$$D_n + E_n = \frac{C_{1n}}{4} \sqrt{2\pi}\, \sigma_T \sqrt{\frac{\pi}{n+1}} \frac{1}{\Omega} e^{-\frac{(\Delta w)^2}{4(n+1)\Omega^2}} \tag{151}$$

and

$$2F_n = \frac{C_{1n}}{4} \sqrt{2\pi\sigma_T} \sqrt{\frac{\pi}{n+1}} \frac{1}{\Omega} e^{-\frac{n}{n+1} \frac{\sigma_T^2}{2} \left[1 + \frac{1}{2\sigma_T^2 \Omega^2 n}\right] (\Delta w)^2} \tag{152}$$

where Ω is the "signal bandwidth" defined by Equation (132).

Substituting into Equation (140) one obtains

$$A_n \leq K_n C_{1n} \sqrt{\frac{2}{n+1}} \frac{1 - e^{-\frac{n}{n+1} \frac{\sigma_T^2}{2} \left[1 + \frac{1}{2\sigma_T^2 \Omega^2 n}\right] (\Delta w)^2}}{1 - e^{-\frac{\sigma_T^2}{4} \left[1 + \frac{1}{2\sigma_T^2 \Omega^2}\right] (\Delta w)^2}} \tag{153}$$

The signal spectrum has maximum slope at $w = \pm \sqrt{2}\,\Omega$. With the substitution $(\Delta w)^2 = 2\Omega^2$, Equation (153) becomes

$$A_n \leq K_n C_{1n} \sqrt{\frac{2}{n+1}} \; \frac{1-e^{-\frac{1}{2(n+1)}} \; e^{-\frac{n}{n+1}\sigma_T^2\Omega^2}}{1-e^{-\frac{1}{2}} \; e^{-\frac{1}{2}\sigma_T^2\Omega^2}} \tag{154}$$

One can readily demonstrate by direct computation that the right side of Equation (154) does not exceed $K_n C_{1n}$ for any value of $\sigma_T^2\Omega^2$ and any $n \geq 3$. As anticipated, the exponential terms [proportional to the ratio of $2F_n/(D_n+E_n)$] are small unless $\Omega \approx \frac{1}{\sigma_T}$. The lowest value of Ω is reached when $K = 0$, in which case $\Omega^2 = \frac{1}{2\sigma_T^2}$ and

$$A_n \leq K_n C_{1n} \sqrt{\frac{2}{n+1}} \tag{155}$$

Thus the bound

$$R \geq 0.89 \tag{156}$$

remains valid for Doppler estimation,[1] at least in this particular example and - from the preceding discussion - probably for most cases of practical interest.

---

[1]Note that we are working with $w_D = 0$. Our conclusions, therefore, refer to clipping losses aside from those due to spectral separation of signal and reverberation.

Appendix A

Consider the problem of detection in a spherically isotropic ambient noise field. Here

$$\rho_{ij}(\tau) = \frac{1}{2\tau_{ih}} \int_{\tau-\tau_{ih}}^{\tau+\tau_{ih}} \rho(\lambda)\, d\lambda \qquad \text{(A-1)}^{1}$$

where $\rho(\lambda)$ is the normalized autocorrelation of the noise and

$$\tau_{ih} = \frac{d_{ih}}{c} \qquad \text{(A-2)}$$

d_{ih} is the distance between the i^{th} and h^{th} hydrophone. If the system processes only a narrow band near the nominal signal frequency, one can write as before

$$\rho(\lambda) = \rho_1(\lambda) \cos w_0 \lambda \qquad \text{(A-3)}$$

Suppose, now, that $\rho_1(\lambda)$ is essentially constant over the interval $\tau - \tau_{ih} \leq \lambda \leq \tau + \tau_{ih}$ for every pair of hydrophones and every value of τ. This is the narrow-band assumption discussed in detail in Section IV.

Then Equation (A-1) becomes

$$\rho_{ij}(\tau) = \frac{\rho_1(\tau)}{2\tau_{ih}} \int_{\tau-\tau_{ih}}^{\tau+\tau_{ih}} \cos w_0 \sigma d\sigma$$

$$= \frac{\sin w_0 \tau_{ih}}{w_0 \tau_{ih}} \rho_1(\tau) \cos w_0 \tau \qquad \text{(A-4)}$$

---

[1] R. A. McDonald, P. M. Schultheiss, F. B. Tuteur, T. Usher. Processing of Data from Sonar Systems, Vol. I, A-1, Equation 3, September 1963.

Returning to Equation (20) and expanding the $\sin^{-1}$ term as in Equation (22), one obtains the following equivalent of the latter equation

$$\frac{1}{R^2} = 1 + K_3 \frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}^3 (t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij} (t-\lambda-\tau_i+\tau_j)} +$$

$$+ K_5 \frac{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij}^5 (t-\lambda-\tau_i+\tau_j)}{\int_0^T dt\, s(t) \int_0^T d\lambda\, s(\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \rho_{ij} (t-\lambda-\tau_i+\tau_j)} + \ldots \qquad \text{(A-5)}$$

where K_n is given by Equation (25).

Equation (A-5) is of the same form as Equation (32). Hence, if one again postulates a signal of the form (31) one can pursue an argument entirely parallel to Equations (33) - (36). The typical term of Equation (A-5) now becomes

$$A_n = K_n C_{1n} \frac{\int_0^T dt s_1(t) \int_0^T d\lambda s_1(\lambda) \rho_1^n (t-\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \left[\frac{\sin w_0 \tau_{ij}}{w_0 \tau_{ij}} \right]^n \cos[w_0(\tau_i-\tau_j)+\phi(t)-\phi(\lambda)]}{\int_0^T dt s_1(t) \int_0^T d\lambda s_1(\lambda) \rho_1 (t-\lambda) \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\sin w_0 \tau_{ij}}{w_0 \tau_{ij}} \cos[w_0(\tau_i-\tau_j)+\phi(t)-\phi(\lambda)]} \qquad \text{(A-6)}$$

In deriving this equation $\rho_1(t-\lambda-\tau_i+\tau_j)$ has been replaced with

$\rho_1(t-\lambda)$, invoking once more the narrow-band assumption, that the correlation time is large compared with sound travel time across the array. The definition of C_{1n} is given in Equation (35).

If we consider an array confined to a plane and a target producing a plane wave parallel to the plane of the array (remote broadside target), then $\tau_i = \tau_j$ for all i and j and Equation (A-6) reduces to

$$A_n = K_n C_{1n} \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{\sin w_0\tau_{ij}}{w_0\tau_{ij}}\right)^n}{\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\sin w_0\tau_{ij}}{w_0\tau_{ij}}} \times \frac{\int_0^T dt s_1(t)\int_0^T d\lambda s_1(\lambda)\rho_1^n(t-\lambda)\cos[\phi(t)-\phi(\lambda)]}{\int_0^T dt s_1(t)\int_0^T d\lambda s_1(\lambda)\rho_1(t-\lambda)\cos[\phi(t)-\phi(\lambda)]} \tag{A-7}$$

The ratio of integrals is identical with that in Equation (38) and therefore has an upper bound of unity. This upper bound can be approached arbitrarily closely, for instance, by choosing $\phi(t) = 0$ (no frequency modulation) and taking $\rho_1(\tau) \doteq 1$ for $|\tau| \leq T$ (narrow-band noise). Thus

$$A_n \leq K_n C_{1n} \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{\sin w_0\tau_{ij}}{w_0\tau_{ij}}\right)^n}{\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\sin w_0\tau_{ij}}{\sin w_0\tau_{ij}}} \tag{A-8}$$

The upper bound can be approached as closely as desired by proper choice of $\rho_1(\tau)$.

Consider now a planar array constructed from equilateral triangles. Figure A1 shows the most elementary version of such an array, with only

seven hydrophones. For it A_n assumes a maximum value of $1.72\ K_n C_{1n}$ at $\omega_0 \tau_{12} = 5.0$.

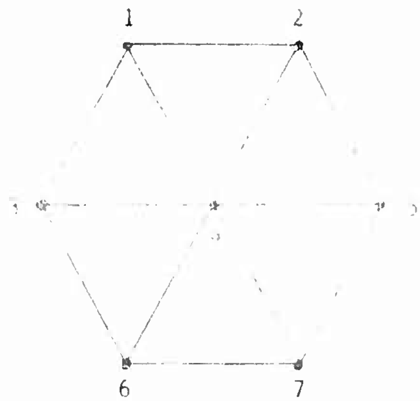


Figure A1

For the next larger regular hexagon (Figure A2, nineteen hydrophones), the maximum A_n is $2.53\ K_n C_{1n}$. Going to a still larger regular hexagon (seven phones along a diameter for a total of thirty-seven phones), one arrives at a maximum A_n of $3.02\ K_n C_{1n}$. In each case A_5 , A_7 .... are only slightly larger than A_3 because only the terms $i = j$ contribute significantly to the numerator of Equation (A-8). Using $A_n = 3.02\ K_n C_{1n}$ and following the procedure of Equations (39) - (43) one obtains

$$R \approx 0.74 \tag{A-9}$$

Thus, the clipping loss has increased to [illegible].6 db (from 1 db for uncorrelated phones).[1] No attempt was made to evaluate more complex array structures, but it appears not unreasonable that still larger clipping losses might be encountered. The physical reasons for the observed phenomenon are clear: [illegible] phones is carefully

[1] In terms of equivalent input signal to noise ratio.

selected to yield high negative noise correlation between adjacent phones. This reduces the effective noise power in the unclipped case, but is ineffective in the clipped case because the generated harmonics do no retain the same phase relations as the fundamentals.

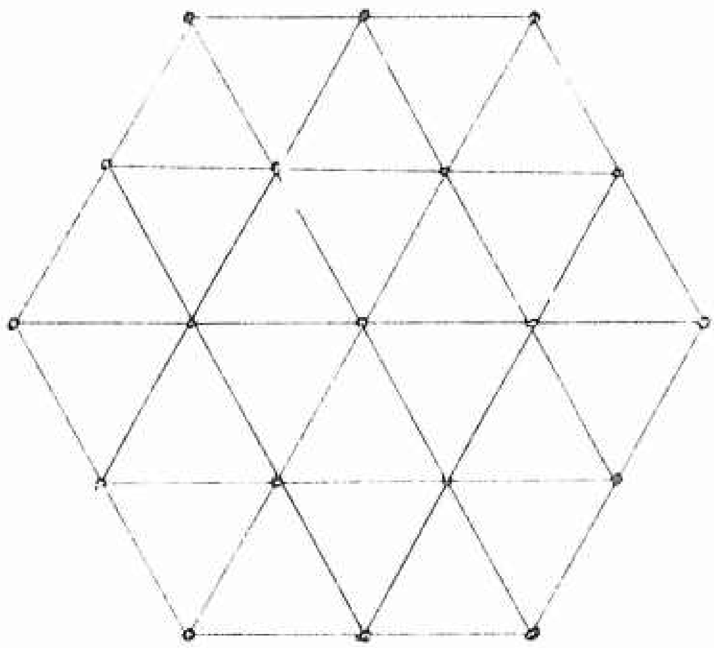


Figure A2

Appendix B

In order to exhibit the influence of the transmitter pattern on the narrow-band approximation, return to Equation (67) and evaluate R_{ij} at the point $\tau - \tau_i + \tau_j$. Using Equation (84)

$$R_{ij}(\tau-\tau_i+\tau_j) = \frac{1}{2} E\left\{ \sum_{\ell} \frac{b_\ell^2}{t_\ell^2\left(t_\ell-\alpha\frac{d_{ij}}{c}\right)^2} s_1(t-t_\ell) s_1\left[t-t_\ell+\frac{d_{ij}}{c}(\alpha-\alpha_{ij})+\tau\right] \times \right.$$

$$\left. \times \cos\left[w_0\tau_0 + w_0\frac{d_{ij}}{c}(\alpha-\alpha_{ij}) + \phi\left[t-t_\ell+\tau+\frac{d_{ij}}{c}(\alpha-\alpha_{ij})\right] - \phi(t-t_\ell)\right]\right\} \qquad \text{(B-1)}$$

Let

$$x = t-t_\ell + \frac{d_{ij}}{c}(\alpha-\alpha_{ij}) + \frac{\tau}{2} \qquad \text{(B-2)}$$

Then, using Equations (68) and (69) one obtains

$$R_{ij}(\tau-\tau_i+\tau_j) = \frac{c^3}{16V}\sum_{\ell}\frac{b_\ell^2}{t_0^2}\int_0^{2\pi} d\phi_\ell \int_0^{\pi} d\theta_\ell\, g(\theta_\ell-\theta_0, \phi_\ell-\phi_0) \times$$

$$\cdot \int_{-\infty}^{\infty} dx\, s_1\left[x-\frac{\tau}{2}-\frac{d_{ij}}{c}(\alpha-\alpha_{ij})\right] s_1\left(x+\frac{\tau}{2}\right)\cos\left\{w_0\tau+w_0\frac{d_{ij}}{c}(\alpha-\alpha_{ij})+\phi\left(x+\frac{\tau}{2}\right)-\phi\left[x-\frac{\tau}{2}-\frac{d_{ij}}{c}(\alpha-\alpha_{ij})\right]\right\} \qquad \text{(B-3)}$$

For narrow beams one obtains from Equation (81)

$$\alpha - \alpha_{ij} = \beta_{ij}(\theta_\ell-\theta_0) + \gamma_{ij}(\phi_\ell-\phi_0) \qquad \text{(B-4)}$$

where β_{ij} and γ_{ij} are defined by Equations (83) and (84). It is a simple matter to demonstrate that

$$\beta_{ij} \ll 1 \quad \text{and} \quad \gamma_{ij} \ll 1 \qquad \text{(B-5)}$$

Hence,

$$\max |\alpha - \alpha_{ij}| \le \max_{\theta_\ell} |\theta_\ell - \theta_0| + \max_{\phi_\ell} |\phi_\ell - \phi_0| \tag{B-6}$$

Approximation 1) [p. 27] is at least as good as before. In place of approximation 2) one now has

$$\phi\left[x - \frac{\tau}{2} - \frac{d_{ij}}{c}(\alpha - \alpha_{ij})\right] \approx \phi(x - \frac{\tau}{2}) - \phi'(x - \frac{\tau}{2})\frac{d_{ij}}{c}(\alpha - \alpha_{ij})$$

$$\cong \phi(x - \frac{\tau}{2}) \tag{B-7}$$

Thus, one makes the implicit assumption

$$\left[\max \phi'(x - \frac{\tau}{2})\right] \frac{d_{ij}}{c} \left[\max(\alpha - \alpha_{ij})\right] << 1 \tag{B-8}$$

Using Equation (B-6) this becomes

$$\left[\max \phi'(x - \frac{\tau}{2})\right] \frac{d_{ij}}{c} \left[\max_{\theta_\ell} |\theta_\ell - \theta_0| + \max_{\phi_\ell} |\phi_\ell - \phi_0|\right] << 1 \tag{B-9}$$

The maximizations in θ_ℓ and ϕ_ℓ extend over the width of the transmitter pattern. For a maximum frequency deviation of 50 cps and a maximum distance between hydrophones of 20 ft., Equation (B-9) reads

$$0.4\pi \text{ (sum of } \theta \text{ and } \phi \text{ pattern half-widths)} << 1 \tag{B-10}$$

The pattern half-widths are measured in radians. Hence, the patterns need not be very narrow before Equation (B-9) becomes substantially less restrictive than its equivalent in Section IV.

Appendix C

Single Frequency Pattern Functions

Consider a transmitting array consisting of N omnidirectional point transducers in an arbitrary (but specified) geometrical arrangement. We wish to find the angular distribution of power in the far field. The basic geometry is shown in Figure C1 and is basically the same as that of Figure 2.

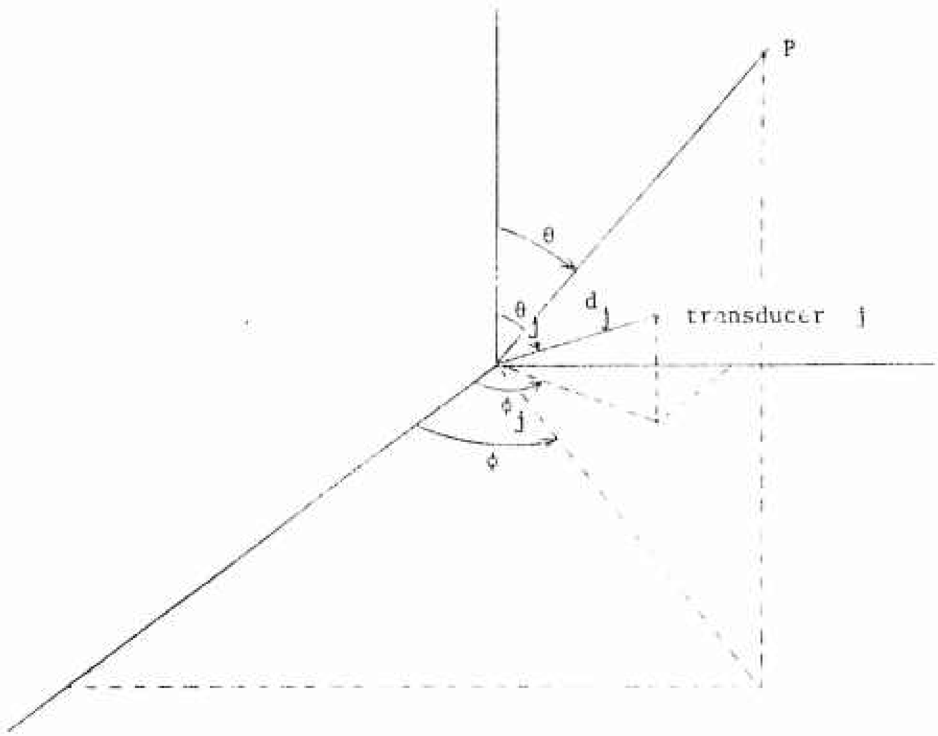


Figure C1

Ignoring spherical spreading losses [which are not part of the coefficients a_ℓ in Equations (61) and (62)] the power at point P is

$$\left| \sum_{j=1}^{N} \; [illegible] \; i\omega(t-t_{j} \; [illegible] \right| \tag{C-1}$$

where t_j is the delay of the signal applied to the j^{th} transducer (steering) and τ_j is the sound travel time from the j^{th} transducer to point P . The pattern function g( ) measures the angular dependence of power and is therefore directly proportional to (C-1).

From Equation (65) one obtains

$$\tau_j = t_0 - \frac{d_j}{c}\left[\sin\theta \sin\theta_j \cos(\phi-\phi_j) + \cos\theta \cos\theta_j\right] \tag{C-2}$$

where t_0 is the sound travel time from the origin to point P and d_j is the distance of the j^{th} transducer from the origin. If the array is steered in the direction of (θ_0, ϕ_0) then

$$t_j = \frac{d_j}{c}\left[\sin\theta_0 \sin\theta_j \cos(\phi_0 - \phi_j) + \cos\theta_0 \cos\theta_j\right] \tag{C-3}$$

If the pattern is confined to small angular deviations from (θ_0, ϕ_0) one can approximate (C-2) by the first terms of a Taylor series

$$\tau_j \cong t_0 - \frac{d_j}{c}\left[\alpha_{0j} + \beta_{0j}(\theta - \theta_0) + \gamma_{0j}(\phi - \phi_0)\right] \tag{C-4}$$

where α_{0j}, β_{0j}, γ_{0j} are analogous to the parameters α_{ij}, β_{ij}, γ_{ij} defined in Equations (80) - (82). From inspection of Equations (80) and (C-3)

$$t_j = \frac{d_j}{c}\alpha_0 \tag{C-5}$$

Now, substituting Equations (C-4) and (C-5) into (C-1)

$$g(\theta-\theta_0, \phi-\phi_0) = \left|\sum_{j=1}^{N} e^{iw\left[t-t_0+\beta_{0j}(\theta-\theta_0)+\gamma_{0j}(\phi-\phi_0)\right]}\right|^2$$

$$= \sum_{j=1}^{N}\sum_{k=1}^{N} e^{iw\left[(\beta_{0j}-\beta_{0k})(\theta-\theta_0)+(\gamma_{0j}-\gamma_{0k})(\phi-\phi_0)\right]} \tag{C-6}$$

Hence, the Fourier transform of g( ) is

$$G(u, v) = \sum_{j=1}^{N} \sum_{k=1}^{N} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \; e^{iw\left[(\beta_{0j}-\beta_{0k})x+(\gamma_{0j}-\gamma_{0k})y\right]} e^{-i(ux+vy)}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{N} \int_{-\infty}^{\infty} dx \; e^{ix\left[(\beta_{0j}-\beta_{0k})w-u\right]} \int_{-\infty}^{\infty} dy \; e^{iy\left[(\gamma_{0j}-\gamma_{0k})w-v\right]}$$

$$= 4\pi^2 \sum_{i=1}^{N} \sum_{k=1}^{N} \delta\left[u-w(\beta_{0j}-\beta_{0k})\right] \delta\left[y-w(\gamma_{0j}-\gamma_{0k})\right] \tag{C-7}$$

Thus, G(u, v) is a non-negative function. This result clearly remains true if one considers the pattern function defined by the total power over some frequency band, for the w integral of (C-7) is clearly non-negative.

SOME COMMENTS ON OPTIMUM BEARING ESTIMATION

by

Peter M. Schultheiss

Progress Report No. 32

General Dynamics/Electric Boat Research

(8050-31-55001)

July 1967

DEPARTMENT OF ENGINEERING

AND APPLIED SCIENCE

YALE UNIVERSITY

Summary

The problem of optimum bearing estimation is discussed for the simplest possible case. The receiving array consists of two hydrophones. Signal and noise are stationary Gaussian random processes with zero means and spectra of the same form. The noise at one hydrophone is statistically independent from that at the other. The following results are obtained.

1) If the output signal-to-noise ratio is large enough to permit accurate bearing measurements, the rms bearing error of a simple cross-correlator (with pre-whitening) reaches the Cramer-Rao lower bound for all input signal-to-noise ratios. Hence the cross-correlator is an optimal bearing estimator.

2) In terms of the input signal-to-noise ratio (S/N) the rms bearing error of the cross-correlator varies as $(S/N)^{-1}$ for $(S/N) \ll 1$ and as $(S/N)^{-\frac{1}{2}}$ for $(S/N) \gg 1$. Thus the optimum bearing estimator has characteristics normally associated with a coherent (incoherent) device for high (low) input signal-to-noise ratios.

3) The rms error of the cross-correlator varies linearly with the correlation time (inverse bandwidth) of signal and noise. It varies inversely with the half power of the time-bandwidth product.

4) The rms bearing error of a 2 element split beam tracker is equal to that of the cross-correlator if the phase shift between beams is achieved by a pure differentiator. If a pure 90° phase shift is used in place of the differentiator there is a small loss in performance, equivalent to about 0.6 db of input signal-to-noise ratio.

I. Introduction

The purpose of this note is to correlate and extend certain previously reported results on optimum and suboptimum bearing estimation. The stimulus was provided by a recent paper by Middleton[1] using a very different approach and obtaining, in part, different results.

In the following analysis the problem of optimality is approached by using the Cramér-Rao inequality to set a lower bound on the attainable rms bearing error. Realizability is then established by describing an instrumentation which actually reaches the lower bound. Only the simplest physical situation is considered. Signal and noise are assumed to be stationary Gaussian random processes with zero means and spectra of the same form. The noise is assumed to be statistically independent from hydrophone to hydrophone. The observation time is long compared with the correlation times of signal and noise. In most of the discussion the receiving array consists of a single pair of omnidirectional hydrophones.

---

[1] D. Middleton, "Optimum and Suboptimum Bearing Estimation for Deterministic and Random Signals in Normal Noise Fields," Paper Ef7, Acoustical Society of America, 73rd Meeting, April 1967, New York. For a more detailed treatment of the same [illegible], see also, D. Middleton, "Extraction of Random Acoustic Signals by Receivers with Distributed Elements," Raytheon Company Report (Submarine Signal Division), [illegible].

II. Lower Bound on RMS Bearing Error

If the receiving array consists simply of a pair of omnidirectional hydrophones the problem of bearing estimation is equivalent to the problem of estimating the signal delay between the two hydrophones. Let the true and estimated delay be δ_o and δ^* respectively. For the case of signal and noise processes with the properties of Gaussian white noise limited in frequency to $0 \leq f \leq W$, McDonald[1] has calculated the Cramèr-Rao lower bound on $\overline{(\delta_o-\delta^*)^2}$.

$$\overline{(\delta_o-\delta^*)^2} \geq \frac{(S+N_1)\ (S+N_2) - S^2}{S^2} \ \frac{1}{8W^2} \ \frac{1}{\frac{\pi^2}{3} WT - \log(2WT-1) - 0.5772} \tag{1}$$

S is the signal power, N_1, N_2 the noise power at phones 1 and 2 respectively and T the observation time. If $N_1 = N_2 \equiv N$ and $WT >> 1$, Equation (1) reduces to

$$\overline{(\delta-\delta^*)^2} \geq \frac{3}{8\pi^2 \ W^2 (WT)} \left(2 \frac{N}{S} + \frac{N^2}{S^2}\right) \tag{2}$$

In order to translate Equation (2) into a lower bound on bearing accuracy, consider the geometry of Figure 1. If the target distance r is very large compared with the spacing d between hydrophones, one can write to an excellent approximation

$$\delta = \frac{d}{c} \sin \theta \tag{3}$$

---

[1] A. L. Levesque, R. A. McDonald, [illegible], [illegible] Usher. "Processing of Data from Sonar Systems," Vol. II, 1964, Appendix B.

c is the velocity of sound in water.

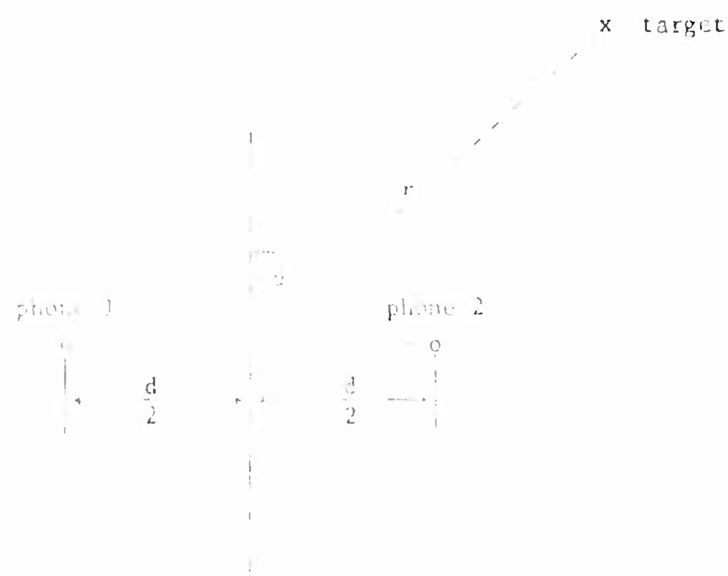


Fig. 1

If the true bearing is θ_o, then the estimation error $\theta^* - \theta_o$ is related to $\delta^* - \delta_o$ by

$$\delta^* - \delta_o = \frac{d}{c}(\sin\theta^* - \sin\theta_o) = 2\frac{d}{c}\sin\tfrac{1}{2}(\theta^* - \theta_o)\cos\tfrac{1}{2}(\theta^* + \theta_o) \tag{4}$$

In practice, one is concerned with the numerical value of the rms tracking error only if it is quite small, i.e., if $\theta^* - \theta_o$ does not exceed a small fraction of a radian with any significant probability. In that case, one can approximate

[illegible] (5)

and, except for θ_o extremely close to $\frac{\pi}{2}$,

$$\cos \tfrac{1}{2}(\theta^* + \theta_o) \simeq \cos \theta_o \tag{6}$$

Substituting these approximations in Equation (4) one obtains

$$\delta^* - \delta_o \cong (\theta^* - \theta_o) \frac{d}{c} \cos \theta_o \tag{7}$$

Hence, the standard deviation of $(\theta^* - \theta_o)$ has the lower bound

$$D(\theta^*) \equiv \sqrt{\overline{(\theta^* - \theta_o)^2}} \geq \frac{\sqrt{3}\, c}{2\sqrt{2}\, \pi W \sqrt{WT}\, d \cos \theta_o} \sqrt{2 \frac{N}{S} + \frac{N^2}{S^2}} \tag{8}$$

This result generalizes in trivial fashion to non-white signal and noise spectra, as long as the ratio of the two spectral functions is a constant[1] over the processed frequency band $0 \leq f \leq W$. For then one can regard the output of each hydrophone as prewhitened by an appropriate linear filter, prior to further processing. The (invertible) linear filtering operation can clearly not have any effect on the minimum attainable[2] mean square error. Thus, Equation (8) remains valid.

---

[1]In practice this condition is often at least approximately true, because signal and noise spectra are shaped by the same hydrophone and receiver characteristics.

[2]This argument does not prove formally that the Cramér-Rao lower bound might not be lower for non-white spectra. However, it does prove that any such lower value cannot be realizable. The optimum estimator after prewhitening cannot achieve a performance better than the right side of Equation (8). But the optimum estimator would include linear filters to shape the spectra to the most desirable form. Therefore, no realizable estimator working with the initial spectra can improve on the lower bound of Equation (8).

III. An Optimal Instrumentation: The Cross-correlator

When only two hydrophones are available an obvious instrumentation for delay measurement (and hence bearing measurement) is the simple cross-correlator shown in Figure 2.

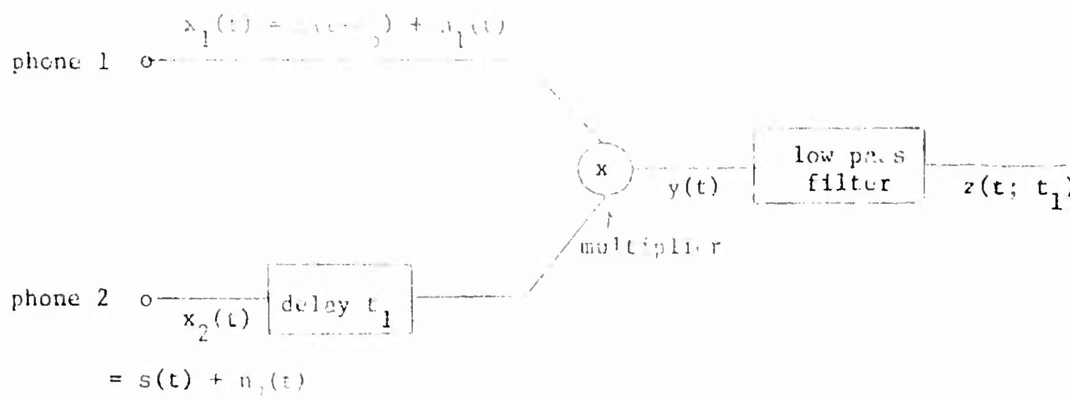


Figure 2

The system output is the short-time cross-correlation (averaged over the smoothing time of the filter) of the two hydrophone outputs, delayed relative to each other by t_1. At any given time t, $z(t; t_1)$ is computed for all possible values of t_1. The value t_1^* of t_1 which maximizes $z(t; t_1)$ is designated as the instantaneous estimate of signal delay. The bearing estimate is then obtained from an obvious analogue of Equation (3).

$$t_1^* = [illegible] \quad (9)$$

The maximum of $z(t;\, t_1)$ occurs at a point where

$$\frac{\partial z}{\partial t_1} = 0 \tag{10}$$

Thus the rms error in t_1 is equal to the rms fluctuation of the null of this derivative. This fluctuation is given by

$$\sigma_{t_1} = \left. \frac{D\left(\frac{\partial z}{\partial t_1}\right)}{\left|\frac{\partial^2 z}{\partial t_1^2}\right|} \right|_{t_1 = \delta_o} \tag{11}$$[1]

D ( ) denotes the standard deviation of the bracketed quantity.

The output $y(t)$ of the multiplier in Figure 2 is given by

$$y(t) = \left[s(t-\delta_o) + n_1(t)\right] \left[s(t-t_1) + n_2(t-t_1)\right] \tag{12}$$

If the noises at the two receivers are statistically independent of each other and of the signal (and have zero means)

$$\overline{z} = \overline{y} = S\, \rho_s(t_1-\delta_o) \tag{13}$$

$\rho_s(\tau)$ is the normalized autocorrelation function of the signal.

---

[1] We are assuming that $\partial z/\partial t_1$ has only one zero in the neighborhood of the correct value, $t_1 = \delta_o$. For signal-to-noise ratios high enough to permit tracking with reasonable accuracy this should be a good assumption.

In evaluating slope and standard deviation at $t_1 = \delta_o$, we imply further that the null of $\partial z/\partial t_1$ fluctuates by less than the correlation time of the signal. Examination of Equation (36) reveals that this is true when the output signal-to-noise ratio is large compared with unity. This is the only condition under which the bearing accuracy problem has much practical interest.

Differentiating Equation (13) twice with respect to t_1 and evaluating at $t_1 = t_0$, one obtains

$$\left.\frac{\partial^2 z}{\partial t_1^2}\right|_{t_1 = t_0} = S\,\phi_{s}''(0) \tag{14}$$

If the lowpass filter in Figure 2 has a weighting function $h(\sigma)$, the output $z(t)$ is given by

$$z(t) = \int_0^{\infty} d\sigma\, h(\sigma)\, y(t-\sigma) \tag{15}$$

hence,

$$\frac{\partial z}{\partial t_1} = \int_0^{\infty} d\sigma\, h(\sigma)\, \frac{\partial y(t-\sigma)}{\partial t_1} \tag{16}$$

We must consider a filter which smoothes the output over the past T seconds. Such a filter has the weighting function

$$h(\sigma) = \begin{cases} \dfrac{1}{T} & 0 \le \sigma \le T \\ 0 & \text{elsewhere} \end{cases} \tag{17}$$

If T is much larger than the correlation time of signal and noise, it is a simple matter to demonstrate [Report No. 10, Equation (9)] that

$$\sigma^2\left(\frac{\partial z}{\partial t_1}\right) = \frac{1}{T} \int_{-\infty}^{\infty} \left[R_{\frac{\partial y}{\partial t_1}}(\tau) - R_{\frac{\partial y}{\partial t_1}}(\infty)\right] d\tau \tag{18}$$

$R_{\frac{\partial y}{\partial t_1}}(\tau)$ is the autocorrelation function of $\frac{\partial y}{\partial t_1}$.

From Equation (12)

$$\frac{\partial y}{\partial t_1} = - \left[s(t-\delta_o) + n_1(t)\right] \left[s'(t-t_1) + n_2' (t-t_1)\right] \tag{19}$$

Hence,

$$R_{\frac{\partial y}{\partial t_1}}(\tau) =$$

$$E\{\left[s(t-\delta_o)+n_1(t)\right]\left[s(t+\tau-\delta_o)+n_1(t+\tau)\right]\left[s'(t-t_1)+n_2'(t-t_1)\right]\left[s'(t+\tau-t_1)+n_2'(t+\tau-t_1)\right]\}$$

(20)[1]

Twelve of the sixteen terms in this average vanish immediately, because they contain one of the noise components only once. The remaining terms are

$$R_{\frac{\partial y}{\partial t_1}}(\tau) = E\{s(t-\delta_o)s(t+\tau-\delta_o)s'(t-t_1)s'(t+\tau-t_1)\}$$
$$+ SN\rho_s(\tau)\rho_n'(\tau) + SN\rho_s'(\tau)\,\rho_n(\tau) + N^2\,\rho_n(\tau)\,\rho_n'(\tau) \tag{21}$$

$\rho_x(\tau)$ designates the normalized autocorrelation function of the random process $x(t)$. Both noise processes are assumed to have the same autocorrelation function $\rho_n(\tau)$.

Since the signal process is Gaussian, the first term of Equation (21) can be expressed in terms of second order moments. Thus

$$R_{\frac{\partial y}{\partial t_1}}(\tau) = S^2\rho_s(\tau)\rho_s'(\tau) + R^2_{ss'}(\delta_o-t_1) + R_{ss'}(\delta_o-t_1+\tau)\,R_{ss'}(\delta_o-t_1-\tau)$$
$$+ SN\left[\rho_s(\tau)\,\rho_n'(\tau) + \rho_s'(\tau)\,\rho_n(\tau)\right] + N^2\,\rho_n(\tau)\,\rho_n'(\tau) \tag{22}$$

---

[1] E{ } stands for the expectation of the bracketed quantity.

Here $R_{xx}(\tau)$ is the (unnormalized) cross-correlation of $x(t)$ and $y(t)$.

There remains the computation of the various correlation functions

$$\rho_{s'}(\tau) = \frac{E\{s'(t)s'(t+\tau)\}}{S} = \lim_{\Delta t \to 0} E\left\{\frac{[s(t+\Delta t) - s(t)]\,[s(t+\tau-\Delta t) - s(t+\tau)]}{S\,(\Delta t)^2}\right\}$$

$$= \lim_{\Delta t \to 0} \frac{2\rho_s(\tau) - \rho_s(\tau+\Delta t) - \rho_s(\tau-\Delta t)}{(\Delta t)^2} \tag{23}$$

Now expanding the last two terms into Taylor series about the point τ :

$$\rho_{s'}(\tau) = \lim_{\Delta t \to 0} \frac{1}{(\Delta t)^2}\Big\{2\rho_s(\tau) - [\rho_s(\tau) + \rho_s'(\tau)\Delta\tau + \rho_s''(\tau)\frac{(\Delta t)^2}{2} + o(\Delta t)^2]$$
$$- [\rho_s(\tau) - \rho_s'(\tau)\Delta t + \rho_s''(\tau)\frac{(\Delta t)^2}{2} + o(\Delta t)^2]\Big\} \tag{24}$$

The notation $o(\Delta t)^2$ indicates that the remainder term in question approaches zero faster than $(\Delta t)^2$ as $\Delta t \to 0$.

Now carrying out the limiting operation

$$\rho_{s'}(\tau) = -\rho_s''(\tau) \tag{25}$$

By completely analogous computations one obtains

$$\rho_{n'}(\tau) = -\rho_n''(\tau) \tag{26}$$

and

$$R_{ss'}(\tau) = S\rho_s'(\tau) = -R_{ss'}(-\tau) \tag{27}$$

In the absence of DC and other deterministic components of signal

and noise, all of the correlation functions and their derivatives tend to zero as $\tau \to \infty$. Hence from Equation (22)

$$R_{\frac{\partial y}{\partial t_1}}(\infty) = R^2_{ss'}(\delta_o - t_1) = S\rho_s'(\delta_o - t_1) \tag{28}$$

Now substituting Equations (22) and (25) - (28) into Equation (18)

$$D^2\left(\frac{\partial z}{\partial t_1}\right) = -\frac{1}{T}\int_{-\infty}^{\infty} d\tau \left\{ S^2\left[\rho_s(\tau)\rho_s''(\tau) + \rho_s'(\tau+\delta_o-t_1)\,\rho_s'(\tau-\delta_o+t_1)\right]\right.$$
$$\left. + SN\left[\rho_s(\tau)\rho_n''(\tau) + \rho_n(\tau)\rho_s''(\tau)\right] + N^2\rho_n(\tau)\rho_n''(\tau)\right\} \tag{29}$$

Using Parseval's theorem and the real translation theorem

$$D^2\left(\frac{\partial z}{\partial t_1}\right) = \frac{2\pi}{T}\int_{-\infty}^{\infty} dw\, w^2 \left\{ S^2 g_s{}^2(w)\left[1-e^{2jw(\delta_o-t_1)}\right] + 2\,SN\, g_s(w)g_n(w) + N^2 g_n{}^2(w)\right\} \tag{30}$$

where $g_s(w)$ and $g_n(w)$ are the normalized spectral functions of signal and noise respectively. Note that the first term

$$\frac{2\pi}{T}\int_{-\infty}^{\infty} dw\, w^2\, S^2 g_s(w)\left[1-e^{2jw(\delta_o-t_1)}\right] = \frac{2\pi}{T} S^2 \int_{-\infty}^{\infty} dw\, w^2 g_s(w)\left[1-\cos 2w(\delta_o-t_1)\right] \tag{31}$$

exhibits the effect on the output variance of a misadjustment in t_1. For $t_1 = \delta_o$ this term vanishes and one obtains

$$D^2\left(\frac{\partial z}{\partial t_1}\right)\Bigg|_{t_1=\delta_o} = \frac{2\pi}{T}\int_{-\infty}^{\infty} dw\, w^2\left[2\,SN\, g_s(w)g_n(w) + N^2 g_n(w)\right] \tag{32}$$

Noting that

$$\rho_s''(0) = -\int_{-\infty}^{\infty} dw\, w^2 g_s(w) \tag{33}$$

we can now substitute Equations (14) and (32) into Equation (11) to obtain

$$\sigma_{t_1} = \frac{\sqrt{\frac{2\pi}{T}\int_{-\infty}^{\infty} dw\, w^2 \left[2\, SN\, g_s(w) g_n(w) + N^2\, g_n^2(w)\right]}}{S\int_{-\infty}^{\infty} dw\, w^2 g_s(w)} \tag{34}$$

Consider now the white spectra

$$g_s(w) = g_n(w) = \begin{cases} \frac{1}{4\pi W}, & |w| \le 2\pi W \\ 0, & |w| > 2\pi W \end{cases} \tag{35}$$

Straightforward evaluation of the integrals in Equation (34) yields

$$\sigma_{t_1} = \frac{1}{2\sqrt{2}\,\pi W\sqrt{WT}}\sqrt{2\frac{N}{S} + \frac{N^2}{S^2}} \tag{36}$$

which is identical with the square root of the right side of Equation (2). Translation of delay error into bearing error proceeds as in Equations (3) - (8), yielding the result

$$D(\theta^*) = \frac{\sqrt{3}\, c}{2\sqrt{2}\,\pi W\sqrt{WT}\, d\cos\theta_0}\sqrt{2\frac{N}{S} + \frac{N^2}{S^2}}$$

$$= \frac{\sqrt{3}\, c}{2\sqrt{2}\,\pi W\sqrt{WT}\, d\cos\theta_0} \cdot \frac{1}{\frac{\frac{S}{N}}{\sqrt{1+2\frac{S}{N}}}} \tag{37}$$

Thus the instrumentation of Figure 2 actually attains the Cramèr-Rao lower bound. The value of t_1 which maximizes $z(t, t_1)$ is therefore the optimum bearing estimate. It is obvious by inspection that it is an unbiased estimate, as demanded by the formulation of the Cramèr-Rao inequality used in deriving Equation (2).

While the last part of the argument has been carried out for white signal and noise, a trivial modification of Figure 2 generalizes the result to Gaussian signals and noises with arbitrary spectra, as long as the ratio of signal and noise spectra is a constant over the processed frequency range. In such cases one simply inserts a prewhitening filter after each hydrophone, thus reducing the problem to the one just treated.

It is interesting to observe that the rms error of the optimum tracker varies as the inverse first power of S/N for small input signal-to-noise ratios (the typical behavior for incoherent processing) whereas, for large input signal-to-noise ratios it varies as the inverse half power of S/N (the typical behavior for coherent processing). For large input signal-to-noise ratios one can reason, of course, that the waveshape at each hydrophone is a good approximation of the signal waveshape. Thus, one has something akin to knowledge of the signal waveshape, the distinguishing characteristic of coherent operation.

IV. Conventional Split Beam Trackers

An instrumentation frequently used in bearing estimation is the split beam tracker, an elementary version of which is shown in Figure 3.

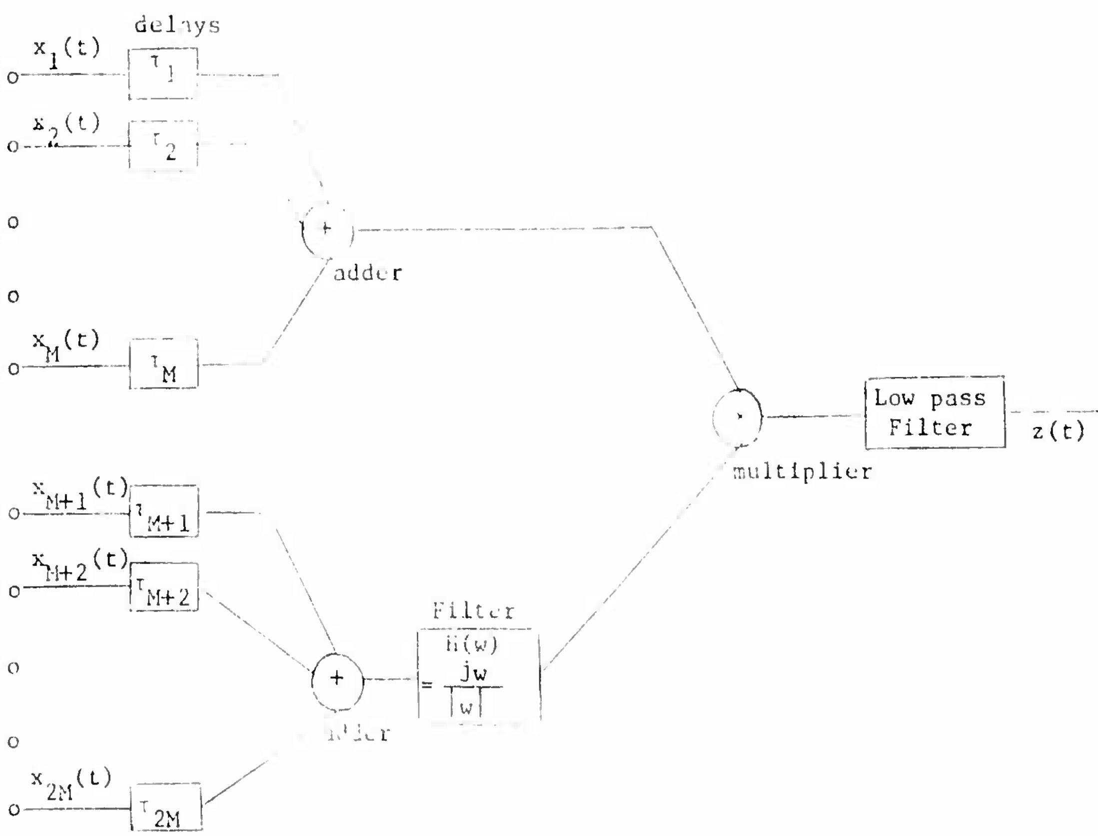


Figure 3

This configuration has been analyzed in detail in Report No. 29. Omitting the interference (the quantity of primary interest in Report No. 29), but retaining the signal dependent terms in the fluctuation, one obtains from Equation (55), (Report No. 29), after a few steps of computation

$$\sigma_z^2 \Big|_{\text{on target}} = \frac{2\pi}{T} \int_{-\infty}^{\infty} dw \left\{ \left|M^2 S g_s(w) + MN g_n(w)\right|^2 - \left|M^2 S g_s(w)\right|^2 \right\} \tag{38}$$

σ_z^2 is the output variance and M is the number of hydrophones in each half of the array. All other symbols retain their earlier meaning.

Using again the white spectra of Equation (35), one obtains, after evaluation of the integrals

$$\sigma_z^2 = \frac{M^2 N^2}{2TW} \left(1 + 2M \frac{S}{N}\right) \tag{39}$$

From Report No. 29, Equation (65)

$$\frac{\partial \bar{z}}{\partial \theta}\Big|_{\text{on target}} = \pi SW M^3 \frac{d}{c} \cos \theta_o \tag{40}$$

Hence,

$$D(\theta) = \frac{\sigma_z}{\frac{\partial \bar{z}}{\partial \theta}}\Bigg|_{\text{on target}} = \frac{c}{\sqrt{2}\, \pi\, W \sqrt{TW}\, M^2\, d \cos \theta_o} \cdot \frac{1}{\frac{\frac{S}{N}}{\sqrt{1+2M\frac{S}{N}}}} \tag{41}$$

For direct comparison with the optimum bearing estimator, consider M = 1 (array consisting of two hydrophones) and take the ratio of Equations (37) and (41)

$$\frac{D(\theta^*)}{D(\theta)} = \frac{\sqrt{3}}{2} \tag{42}$$

Thus the split beam tracker performs almost optionally: The ratio of rms errors is 0.87, equivalent to about 0.6 db of input signal-to-noice ratio.

If in Figure 3 one replaces the pure phase shift

$$H(jw) = \frac{jw}{|w|} \tag{43}$$

with pure differentiator

$$H(jw) = jw \tag{44}$$

Equations (39) and (40) assume the form

$$\sigma_z^2 \Big|_{\text{on target}} = \frac{2\pi^2 W}{3T} \left| 2SN\, M^3 + N^2 M^2 \right| \tag{45}$$

and

$$\frac{\partial \bar{z}}{\partial \theta} \Big|_{\text{on target}} = \frac{4\pi^2 W^2}{3}\, S\, M^3\, \frac{d}{c} \cos\theta_o \tag{46}$$

hence, for $M = 1$

$$D(\theta) = \frac{\sqrt{3}\, c}{2\sqrt{2}\, \pi\, W \sqrt{WT}\, d \cos\theta_o} \cdot \frac{\dfrac{1}{\frac{S}{N}}}{1 + 2\dfrac{S}{N}} \tag{47}$$

This is identical with Equation (37) and with the Cramer-Rao lower bound. Thus, at least in the simple case under discussion, the conventional split beam tracker actually achieves the optimum possible performance.

THE EFFECT OF NOISE ANISOTROPY ON DETECTABILITY IN AN OPTIMUM ARRAY PROCESSOR

By

Franz B. Tuteur

Progress Report No. 33

General Dynamics/Electric Boat Research

September 1967

DEPARTMENT OF ENGINEERING
AND APPLIED SCIENCE

YALE UNIVERSITY

[illegible] Introduction

The object of this report is to investigate the effect of localized noise sources on the performance of the [illegible] (likelihood-ratio) detector.

In a previous progress [illegible] expressions for the performance loss of likelihood-ratio [illegible] the [illegible] isotropic part and a component [illegible] obtained. In the present report these results are [illegible] in an attempt to also get some estimate of the performance loss caused by anisotropy sources that are not strongly localized. Such sources can, presumably, be represented by a large number of closely spaced point sources.

The notation used follows that of [illegible] et. al. [2], since this permits important simplifications over the notation used previously. For the sake of completeness the expressions for the likelihood-ratio processor are rederived here and a brief comparison between the old and new notation is given.

Unfortunately, in spite of the simplified notation the final result is still difficult to evaluate. An approximate result showing that the performance loss due to point noise sources is equivalent to the loss of one hydrophone per point source can be obtained by elimination of all of the terms causing computational difficulties. However, since these terms also contain all the information about the [illegible] of the noise sources this approximation is fairly [illegible] the number of localized noise sources is small and [illegible] from each other and from the target direction. [illegible] obtained for small anisotropic [illegible] that noise [illegible] distributed [illegible] in

azimuth has approximately the same effect on detectability as a single point source of the same strength located at the center of the distributed source. For large anisotropic-to-isotropic noise ratio a similar result can be demonstrated only if the anisotropic noise source extends only over a much smaller angle.

In order to get some idea of performance loss in cases not covered by these analytical results computations were also performed on the digital computer. Although these computations are not conclusive they indicate that the analytical results for distributed noise sources at low anisotropic-to-isotropic noise ratio can probably be extended to large anisotropic-to-isotropic noise ratio.

II. Basic Analysis

Assume that the array consists of M hydrophones, and that the received signal at the i^{th} hydrophone is $x_i(t)$. Then if the spectrum of $x_i(t)$ is limited to frequencies below W cps, and the x(t) are observed over an interval T, such that WT >> 1, $x_i(t)$ can be expanded in a Fourier series:

$$x_i(t) = \sum_{n=-WT}^{W(T)} x_i(n) e^{j2\pi nt/T} \tag{1}$$

where the $x_i(n)$ are complex Fourier coefficients satisfying $x_i(-n) = x_i^*(n)$ and where the asterisk stands for complex conjugate. Then all the available information about the signals received by the entire array is contained in the set of vectors

$$\underline{X}(n) = \begin{bmatrix} x_1(n) \\ \vdots \\ x_M(n) \end{bmatrix} \tag{2}$$

Following Bryn |3| or Edelblute, Fisk, and Kinnison |2| , we assume that $\underline{X}(n)$ and $\underline{X}(m)$ are statistically independent for $n \neq \pm m$. Suppose that the signal $x_i(t)$ received at the i^{th} hydrophone consists of signal and noise; then we let the signal be given by

$$s_i(t) = \sum_{n=-WT}^{WT} y_i(n)\, e^{j2\pi nt/T} \tag{3}$$

so that the signal at all hydrophones is represented by

$$\underline{Y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_M(n) \end{bmatrix} \tag{4}$$

Here again we assume the $\underline{Y}(n)$ to be independent from $\underline{Y}(m)$ for $n \neq \pm m$.

The optimum detector is known to be the likelihood ratio detector, which

determines presence or absence of a signal by comparing the likelihood ratio

$$LR = \frac{f_S(\underline{X})}{f_N(\underline{X})} \tag{5}$$

to a fixed threshold. Here $f_S(\underline{X})$ is the conditional probability density of the received samples (over all [illegible] and over all frequencies) when signal is assumed to be present; similarly $f_N(\underline{X})$ is the conditional probability density when signal is assumed to be absent. Since $\underline{X}(-n) = \underline{X}^*(n)$, and since $\underline{X}(n)$ and $\underline{X}(m)$ are independent for $n \neq \pm m$

$$LR = \prod_{n=1}^{WT} \frac{f_S\left[\underline{X}(n)\right]}{f_N\left[\underline{X}(n)\right]} \tag{6}$$

We assume now that whether signal is present or not $x(t)$ is a stationary Gaussian random process with zero mean value. The normalized covariance matrix for noise only is

$$\underline{Q}(n) = \frac{1}{N(n)} \left\langle \underline{X}^*(n) \underline{X}^T(n) \right\rangle_N \tag{7}$$

where the superscript T refers to transposition and the symbol $\langle\ \rangle_N$ means ensemble average subject to the noise only hypothesis. $N(n)$ is the average noise power at frequency $2\pi n/T$ rad/sec. Then

$$f_N\left[\underline{X}(n)\right] = C_N(n) \exp\left\{ \frac{-1}{N(n)} \left[\underline{X}^T(n)\, \underline{Q}^{-1}(n)\, \underline{X}^*(n) \right] \right\} \tag{8}$$

where $C_N(n)$ is the normalizing constant [illegible].

Assume that signal and noise are independent, and that the normalized covariance matrix for signal alone is given by

$$\underline{P}(n) = \frac{1}{S(n)} \left\langle \underline{Y}^*(n)\, \underline{Y}^T(n) \right\rangle_S \tag{9}$$

where $S(n)$ is the average [illegible].

If the signal is a plane wave, the elements $y_j(n)$ of $\underline{Y}(n)$ are all delayed replicas of each other; thus

$$y_i(n) = c_i s(n)\, e^{j \frac{2\pi n}{T} \tau_i} \tag{10}$$

where $s(n)$ is the n^{th} Fourier coefficient of the signal wave form; the c_i are weighting factors to take into account that the signal strength or gain at different hydrophones may be different, and τ_i is the delay at the i^{th} hydrophone. The c_i's are conveniently defined in such a way that

$$\langle s^*(n)\, s(n) \rangle = S(n) \tag{11}$$

for all n. Hence

$$\underline{Y}(n) = s(n) \begin{bmatrix} c_1 e^{j \frac{2\pi n \tau_1}{T}} \\ \vdots \\ c_M e^{j \frac{2\pi n \tau_M}{T}} \end{bmatrix} \equiv s(n)\underline{V}(n), \tag{12}$$

and therefore

$$\underline{P}(n) = \frac{\langle s^*(n)\, s(n) \rangle}{S(n)}\, \underline{V}^*(n)\underline{V}^T(n) \tag{13}$$

$$= \underline{V}^*(n)\underline{V}^T(n) \tag{14}$$

$\underline{P}(n)$ is seen to be of rank 1. Because of the independence of signal and noise the covariance matrix of signal and noise together is $N(n)\,\underline{Q}(n) + S(n)\,\underline{P}(n)$. Therefore

$$f_s[\underline{X}(n)] = C_{S+N}(n) \exp\left\{ -\underline{X}^T(n) \left[N(n)\,\underline{Q}(n) + S(n)\,\underline{P}(n) \right]^{-1} \underline{X}^*(n) \right\} \tag{15}$$

and therefore the likelihood ratio is

$$LR = \prod_{n=1}^{WT} \frac{C_{S+N}(n)}{C(n)} \exp\left\{ \underline{X}^T(n) \left[\frac{\underline{Q}^{-1}(n)}{N(n)} - \left[N(n)\,\underline{Q}(n) + S(n)\underline{P}(n) \right]^{-1} \right] \underline{X}^*(n) \right\} \tag{16}$$

Since $\underline{P}(n)$ is of rank 1 the inversion of the second term in the brackets is easily accomplished by using the following identity [4]: For an arbitrary nonsingular

matrix $\underline{A}$ and two vectors $\underline{U}$ and $\underline{V}$

$$(\underline{A} + \underline{U}\underline{V}^T)^{-1} = \underline{A}^{-1} - \frac{\underline{A}^{-1}\,\underline{U}\,\underline{V}^T\,\underline{A}^{-1}}{1 + \underline{V}^T\,\underline{A}^{-1}\,\underline{U}} \tag{17}$$

Using this identity

$$\left[N(n)\,\underline{Q}(n) + S(n)\,\underline{P}(n)\right]^{-1} = \frac{\underline{Q}^{-1}(n)}{N(n)} - \frac{S(n)\,\underline{Q}^{-1}(n)\,\underline{P}(n)\,\underline{Q}^{-1}(n)}{N^2(n)\left[1 + S(n)G_o(n)/N(n)\right]} \tag{18}$$

$$\text{where } G_o(n) = \underline{V}^T(n)\,\underline{Q}^{-1}(n)\,\underline{V}^*(n) \tag{19}$$

As is shown by Edelblute et. al. [3], $G_o(n)$ is the maximum value of the array gain. Using this result one finds that the logarithm of the likelihood ratio is given be

$$\log LR = C + \sum_{n=1}^{WT} \frac{S(n)\,\underline{X}^T(n)\,\underline{Q}^{-1}(n)\,\underline{P}(n)\,\underline{Q}^{-1}(n)\,\underline{X}^*(n)}{N^2(n)\left[1 + S(n)\,G_o(n)/N(n)\right]} \tag{20}$$

where $C = \log \prod_{n=1}^{WT} C_{S+N}(n) / C_N(n)$

Since $\underline{Q}^T(n) = \underline{Q}^*(n)$ the quadratic form appearing in this expression can be written in the form:

$$\underline{X}^T(n)\,\underline{Q}^{-1}(n)\,\underline{P}(n)\,\underline{Q}^{-1}(n)\,\underline{X}^*(n) = \underline{X}^T(n)\,\underline{Q}^{-1}(n)\,\underline{V}^*(n)\,\underline{V}^T(n)\,\underline{Q}^{-1}(n)\,\underline{X}^*(n)$$

$$= \left[\underline{V}^{*T}(n)\,\underline{Q}^{*-1}(n)\,\underline{X}(n)\right]\left[\underline{V}^T(n)\,\underline{Q}^{-1}(n)\,\underline{X}(n)\right]^* = \left|\underline{Z}^T(n)\,\underline{X}(n)\right|^2 \tag{21}$$

and therefore

$$\log LR = C + \sum_{n=1}^{WT} \left|\underline{H}^T(n)\,\underline{X}(n)\right|^2 \tag{22}$$

$$\text{where } \underline{H}(n) = \frac{\sqrt{S(n)}\,\underline{Q}^{-1}(n)\,\underline{V}^*(n)}{N(n)\sqrt{1 + S(n)G_o(n)/N(n)}} \tag{23}$$

This implies that log LR can be obtained from a circuit of the form shown in Fig. 1 where the elements of the filter bank for $f = f_n$ are the elements of the vector $H(n)$.

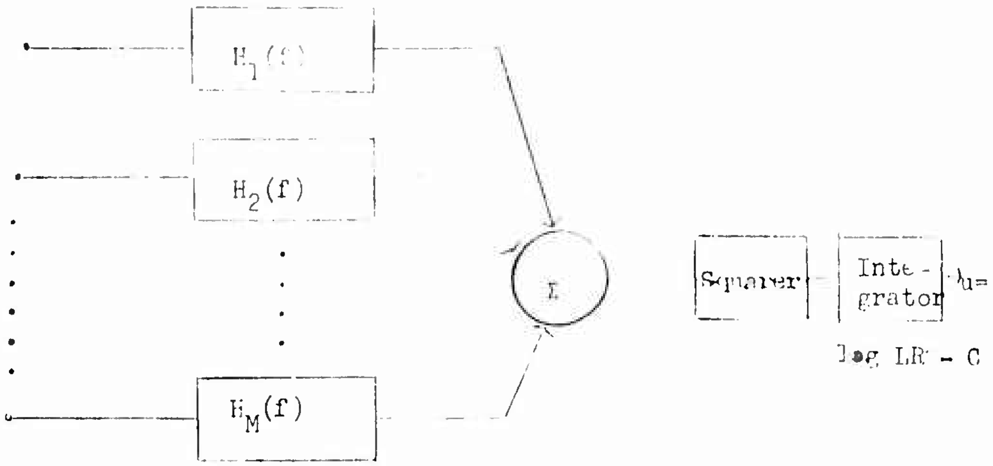


Figure 1 Likelihood Ratio Detector

Except for the scalar multiplier this result is the same as that obtained by Knapp [5] . The difference between the result given here (which is that of Edelblute, et. al.) and Knapp's stems from the fact that Knapp's filter maximizes the output signal-to-noise ratio defined in terms of a signal-plus-noise covariance matrix $\underset{\sim}{C}_X$, rather than producing log LR.

III. Detection Index

In this section it is assumed that the signal-to-noise ratio is small; i.e. $S(n)G_o(n) \ll N(n)$ for all frequencies. In this case one can assume that log LR is approximately a Gaussian random variable and that the performance of the detector is specified by

$$d = \frac{[\text{illegible}]_{S+N} - [\text{illegible}]_{N}}{[\text{illegible}]} \tag{24}$$

where $u = \log LR$, and where $\sigma_N(u)$ is the standard deviation of u under the hypothesis that there is only noise at the input. The computation for d is straight forward; for details see Appendix A. The result is

$$d = \frac{\sum_{n=1}^{WT} K(n) S(n) G_o^2(n)}{\sqrt{\sum_{n=1}^{WT} K^2(n) N^2(n) G_o^2(n)}} \tag{25}$$

$$\text{where } K(n) = \frac{S(n)/N^2(n)}{1 + S(n)\, G_o(n)/N(n)} \approx \frac{S(n)}{N^2(n)} \quad \text{if } S(n)\, G_o(n) \ll N(n) \tag{26}$$

Using the small signal approximation to K(n)

$$d = \sqrt{\sum_{n=1}^{WT} \frac{S^2(n)\, G_o^2(n)}{N^2(n)}} = \sqrt{\sum_{n=1}^{WT} \frac{S^2(n)}{N^2(n)} \left| \underline{V}^T(n)\, \underline{Q}^{-1}(n)\, \underline{V}^*(n) \right|^2} \tag{27}$$

Note that in earlier reports we used

$$d = \sqrt{\frac{1}{2} \sum_{n=1}^{WT} \operatorname{tr} \left[\underline{P}(n)\, \underline{Q}^{-1}(n) \right]^2} \tag{28}$$

This is equivalent to Eq. (27) as can be seen by using the identity

$$\underline{U}^T \underline{A}\, \underline{U}^* = \operatorname{tr} \underline{A}\, \underline{U}^* \underline{U}^T = \operatorname{tr} \underline{U}^* \underline{U}^T \underline{A} \tag{29}$$

where $\underline{U}$ is a vector and $\underline{A}$ is a square matrix. Hence

$$\operatorname{tr} \left[\underline{P}(n)\, \underline{Q}^{-1}(n) \right]^2 = \operatorname{tr} \left\{ \underline{P}(n) \left[\underline{V}^T(n)\, \underline{Q}^{-1}(n)\, \underline{V}^*(n) \right] \underline{V}^T(n)\, \underline{Q}^{-1}(n) \right\} \tag{30}$$

The bracketed [illegible] be factored out; in fact, this

scalar is $G_o(n)$. The remaining terms are also equivalent to $G_o(n)$ by Eqs. (29) and (19); thus $\operatorname{tr}\left[\underline{P}(n)\,\underline{Q}^{-1}(n)\right]^2 = G_o^2(n)$. The factor of 1/2 appearing in Eq (28) does not appear here because of the complex notation that is used. Thus Eqs. (27) and (28) are identical.

As $T \to \infty$ the summation appearing in Eq (27) can be converted to an integral as follows:

$$d = \sqrt{\sum_n \frac{S^2(n)\,G_o^2(n)}{N^2(n)}\,\Delta f} \to \sqrt{T \int_o^W \frac{S^2(f)\,G_o^2(f)}{N^2(f)}\,df} \tag{31}$$

IV. Effect of Directional Interference

Suppose that the noise component of $x_i(t)$ consists of two parts, an isotropic part and an interference part. It is assumed that the interference is generated by R point sources. The r^{th} point source is located at an azimuth angle θ_r, and its spectral density is $I_r(\omega)$; hence the interference power from the r^{th} interference source at the frequency ω_n is $I_r(n)$. The desired target is at the azimuth angle $\theta_o = 0$, and it is assumed that the array is steered in the target direction. The isotropic noise power at the frequency ω_n is $N_o(n)$. The isotropic noise component, the interference sources, and the target signal are all assumed to be mutually independent Gaussian processes with zero mean. Then the total noise power density is given by

$$N(n) = N_o(n) + \sum_{r=1}^{R} I_r(n) \tag{32}$$

and the normalized noise covariance matrix has the form:

$$\underline{Q}(n) = \frac{N_o(n)}{N(n)}\,\underline{Q}_o(n) + \sum_{r=1}^{R} \frac{I_r(n)}{N(n)}\,\underline{V}_r^*(n)\,\underline{V}_r^T(n) \tag{33}$$

$$= \frac{N_o(n)}{N(n)}\left[\underline{Q}_o(n) + \sum_{r=1}^{R} \frac{I_r(n)}{N_o(n)}\,\underline{V}_r^*(n)\,\underline{V}_r^T(n)\right]$$

where [illegible] (n) is the normalized covariance matrix of the isotropic noise component and where each element of the summand results from one of the interference point sources. By direct analogy to Eqs. (12), (13), and (14)

$$\underline{V}_r(n) = [illegible] \qquad (34)$$

where $\tau_i^{(r)}$ is the delay of the plane wave from the r^{th} interference source at the i^{th} hydrophone.

The matrix Q(n) can be inverted by using the following generalization of Eq. (17): if $\underline{A}$ is a nonsingular matrix of dimension M and $\underline{B}$ is a matrix of M rows and R columns then

$$(\underline{A} + \underline{B}\,\underline{B}^T)^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}\,(\underline{I} + \underline{B}^T\underline{A}^{-1}\underline{B})^{-1}\,\underline{B}^T\underline{A}^{-1} \qquad (35)$$

This identity is easily proved by multiplication:

$$(\underline{A} + \underline{B}\,\underline{B}^T)\,[\underline{A}^{-1} - [illegible]\,\underline{B}^T\underline{A}^{-1}]$$

$$= \underline{I} + \underline{B}\,\underline{B}^T\underline{A}^{-1} - \underline{B}\,(\underline{I} + \underline{B}^T\underline{A}^{-1}\underline{B})\,(\underline{I} + \underline{B}^T\underline{A}^{-1}\underline{B})^{-1}\,\underline{B}^T\underline{A}^{-1}$$

$$= \underline{I} + \underline{B}\,\underline{B}^T\underline{A}^{-1} - \underline{B}\,\underline{B}^T\underline{A}^{-1} = \underline{I} \qquad (36)$$

In the present application set

$$\underline{A} = \frac{N(n)}{[illegible]}\,[illegible] \qquad (37)$$

$$[illegible] \sqrt{\frac{B_r [illegible]}{N(n)}}\,\underline{V}_r(n)\Big] \qquad (38)$$

[illegible] $(n)]$

where $K_r = K_r(n) = I_r(n)/N_o(n)$ for $r = 1, \ldots, R$ (39)

Also, to simplify the notation let

$$G_{rs} = G_{rs}(n) = \underline{V}_r^T(n)\, \underline{C}_o^{-1}(n)\, \underline{V}_s^*(n) \tag{40}$$

Note that $G_{rs}(n)$ has the general form of an array gain [illegible] (19). We can call it a "cross array gain". It is clear that

$$G_{rs}^*(n) = G_{sr}(n) \tag{41}$$

In terms of this notation the matrix $\underline{I} + \underline{B}^T \underline{A}^{-1} \underline{B}^*$ of Eq. ([illegible]) becomes

$$\underline{I} + \begin{bmatrix} \sqrt{K_1}\, \underline{V}_1^T \\ \cdots\cdots \\ \sqrt{K_2}\, \underline{V}_2^T \\ \cdots\cdots \\ \cdots\cdots \\ \sqrt{K_R}\, \underline{V}_R^T \end{bmatrix} \underline{C}_o^{-1} \left[\sqrt{K_1}\,\underline{V}_1^* \;\vdots\; \sqrt{K_2}\,\underline{V}_2^* \;\vdots\; \cdots \;\vdots\; \sqrt{K_R}\,\underline{V}_R^* \right]$$

$$= \begin{bmatrix} 1 + K_1 G_{11} & \sqrt{K_1 K_2}\, G_{12} & \cdots\cdots & \sqrt{K_1 K_R}\, G_{1R} \\ \sqrt{K_2 K_1}\, G_{12}^* & 1 + K_2 G_{22} & \cdots\cdots & \sqrt{K_2 K_R}\, G_{2R} \\ \vdots & & & \\ \sqrt{K_R K_1}\, G_{1R}^* & \cdots\cdots & \cdots\cdots & 1 + K_R G_{RR} \end{bmatrix}$$

$$= \underline{I} + \underline{G} \tag{42}$$

by an obvious definition of the "cross-array gain" matrix $\underline{G}$. Note that G is square, of dimensionality R, and Hermitian.

The detection index is given by Eq. ([illegible]) except that for the sake of consistency we rewrite $\underline{V}(n)$ as $\underline{V}_o(n)$. Then

$$d = \sqrt{\sum_{n=1}^{WT} \left[\frac{S(n)}{N(n)}\, \underline{V}_o^T(n) \ldots \right]} \quad \text{[illegible]} \tag{43}$$

$$\sqrt{\sum_{n=1}^{WT} \ldots F(n)} \quad \text{[illegible]} \tag{44}$$

where

$$
\begin{aligned}
F(n) &= \underline{V}_o^T(n)\, \underline{Q}^{-1}(n)\, \underline{V}_o^*(n) \\
&= \underline{V}_o^T \left\{ \underline{Q}_o^{-1} - \underline{Q}_o^{-1} \left[\sqrt{K_1}\,\underline{V}_1^* \;\vdots \cdots\cdots \vdots\; \sqrt{K_R}\,\underline{V}_R^* \right] \left[\underline{I} + \underline{G} \right]^{-1} \begin{bmatrix} \sqrt{K_1}\,\underline{V}_1^T \\ \vdots \\ \sqrt{K_R}\,\underline{V}_R^T \end{bmatrix} \underline{Q}_o^{-1} \right\} \underline{V}_o^* \\
&= G_{oo} - \left[\sqrt{K_1}\,G_{o1} \cdots\cdots \sqrt{K_R}\,G_{oR} \right] \left[\underline{I} + \underline{G} \right]^{-1} \begin{bmatrix} \sqrt{K_1}\,G_{1o} \\ \vdots \\ \sqrt{K_R}\,G_{Ro} \end{bmatrix}
\end{aligned}
\tag{45}
$$

This equation can be written in a more compact form by defining a vector $\underline{g}$ such that

$$
\underline{g}^T = \left[\sqrt{K_1}\,G_{o1} \;\sqrt{K_2}\,G_{o2} \;\cdots\cdots\; \sqrt{K_R}\,G_{oR} \right] \tag{46}
$$

Then, since $G^*_{rs} = G_{rs}$

$$
F(n) = G_{oo} - \underline{g}^T \left[\underline{I} + \underline{G} \right]^{-1} \underline{g} \tag{47}
$$

It is clear that G_{oo} [illegible] the optimum array gain with isotropic noise and that $\underline{g}^T \left[\underline{I} + \underline{G} \right]^{-1} \underline{g}^*$ represents the effect of the interference. In general the evaluation of the interference term in specific instances is difficult for two reasons:

a) The cross-array gains G_{ij} are quadratic forms involving the inverted $\underline{Q}_o(n)$ matrix

b) Even if the G_{ij} are known the [illegible] R matrix $[\underline{I} + \underline{G}]$ must be inverted.

Thus it is necessary either to solve [illegible] by computer or to make approximations permitting an analytic [illegible].

The standard [illegible] in most previous analyses and which eliminates part of the difficulty involved in evaluating the G_{ij} is to assume that there is no correlation between different hydrophones due to the isotropic noise [illegible] has been considered before and is [illegible] in any significant way.

$$G_{oo}(n) = G_{11}(n) = \ldots = G_{RR}(n) = M \tag{48}$$

$$G_{rs}(n) = \sum_{i=1}^{M} e^{j\omega_n(\tau_i^{(r)} - \tau_i^{(s)})} \tag{49}$$

A further small simplification results from the fact that the array is steered on target; this implies that $\tau_i^{(o)} = 0$, and therefore

$$G_{or}(n) = \sum_{i=1}^{M} e^{-j\omega_n \tau_i^{(r)}} \tag{50}$$

Eq.(47) is now explicitly evaluated for a number of simple special cases.

V. Single Point Interference

If there is only a single interference, the Eq.(47) becomes

$$F(n) = G_{oo} - \frac{K_1 \mid G_{o1} \mid^2}{1 + K_1 G_{11}} \tag{51}$$

and with the simplification $\underline{Q}_o(n) = \underline{I}$; this becomes

$$F(n) = M - \frac{K_1 \mid G_{o1} \mid^2}{1 + K_1 M} \tag{52}$$

By Eq (50)

$$\mid G_{o1} \mid^2 = \sum_{i=1}^{M} \sum_{k=1}^{M} e^{j\omega_n(\tau_k - \tau_i)} = \sum_{i=1}^{M} \sum_{k=1}^{M} \cos \omega_n(\tau_k - \tau_i)$$
$$= M + 2 \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \cos \omega_n(\tau_i - \tau_k) \tag{53}$$

where, for simplicity $\tau_i^{(1)} \equiv \tau_i$

Hence the term representing the loss of detectability due to interference in Eq.(47) becomes

$$\frac{K_1 M}{1 + K_1 M} \left[1 + \frac{2}{M} \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \cos \omega_n(\tau_i - \tau_k) \right] \tag{54}$$

This result is essentially identical to Eq.(22) of Ref.[1].

The value of [illegible] sum in Eq. ([illegible]) can be further evaluated for the spe[illegible] [illegible] hydrophones. If such an array is steered [illegible] a direction perpendicular to the array axis [illegible]

$$[illegible] \tag{55}$$

where [illegible] array elements, c is the velocity of sound, and [illegible] from the interference source to the cent[illegible] array axis.

If Eq. ([illegible]) is [illegible], the double sum in ([illegible]) can be replaced by a single sum as shown:

$$\frac{K_1 M}{1 + K_1 [illegible]} \left| 1 + [illegible] (M - k) \cos k [illegible] \right| \tag{56}$$

Substituting this expression [illegible] results in

$$d = \sqrt{\sum_{n=1} [illegible] \frac{[illegible]}{N_o(n)} [illegible] \frac{K_1 [illegible]}{1 + K_1(n [illegible])} \left[1 + \frac{2}{M} \sum_{k=1}^{M-1} (M - k) \cos k\omega_n \tau_o \right] [illegible]^2} \tag{57}$$

where $K_1(n)$ [illegible]

Since our interest [illegible] the effect of [illegible] oscillation term, assume that $S(n)/N_o(n)$ and K_1 [illegible] independent of n [illegible] signal, interference, and isotropic noise [illegible]. [illegible] summation over n can be performed by converting it into an integral, as in Eq. (31); i.e.

$$d = \sqrt{T [illegible] \frac{K_1}{[illegible]} [illegible] (M - k) \cos 2\pi k f \tau_o [illegible]^2 \, df} \tag{58}$$

After [illegible]

$$d^2 = T\omega \frac{S^2}{N_o^2} \left\{ \left(M - \frac{K_1 M}{1 + K_1 M} \right)^2 - \frac{4K_1 M \left[1+K_1 (M-1) \right]}{(1+K_1 M)^2} \sum_{k=1}^{M} (M-k) \frac{\sin 2\pi k W \tau_o}{2\pi k W \tau_o} \right.$$

$$\left. + \frac{2K_1^2}{(1 + K_1 M)^2} \sum_{k}^{M} \sum_{i}^{M} (M-k)(M-i) \left[\frac{\sin 2\pi(k-i)W\tau_o}{2\pi(k-i)W\tau_o} + \frac{\sin 2\pi(k+i)W\tau_o}{2\pi(k+i)W\tau_o} \right] \right\} \tag{59}$$

In most practical cases the maximum frequency processed is such that $2\pi W\tau_o \gg 1$; this is also consistent with the assumption that $\underset{\sim}{Q}_o(n) = I$. Then the summand of the single summation in the second term is small and oscillatory, and the same is true of the last term except when $k=i$; thus the last term is approximately equal to $\dfrac{2K_1^2 M^3}{3(1+K_1 M)^2}$ and

$$d^2 \approx T\omega \frac{S^2}{N_o^2} \left\{ \left[M - \frac{K_1 M}{1+K_1 M} \left(1 - \frac{1}{3} \frac{K_1 M}{1+K_1 M} \right) \right]^2 + O\left(\frac{K_1 M}{1+K_1 M} \right) \right\} \tag{60}$$

where $O(\cdot)$ is a remainder term which is of order $1/M$ relative to the first term. Hence if $K_1 M \gg 1$,

$$d^2 \approx T\omega \frac{S^2}{N_o^2} \left[M - 2/3 \right]^2 \tag{61}$$

which indicates that the effect of the interference is equivalent to the loss of $2/3$ of a hydrophone from the array. This is substantially in agreement with the conclusion of Ref. [1] where it was concluded that the loss under these conditions was equivalent to one hydrophone; this was arrived at by completely neglecting the summation term in Eq. (54). Since the figure of $2/3$ has been derived only for a linear array, and since it probably depends on the array configuration, the figure of one hydrophone is probably a reasonable estimate in general.

VI. Two Point Interferences

With two interferences Eq. ([illegible]) becomes

$$F(n) = G_{oo} - [\sqrt{K_1}G_{o1} \; \sqrt{K_2}G_{o2}] \begin{bmatrix} 1+K_1G_{11} & \sqrt{K_1K_2}G_{12} \\ \sqrt{K_1K_2}G_{12}^* & 1+K_2G_{22} \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{K_1}G_{o1}^* \\ \sqrt{K_2}G_{o2}^* \end{bmatrix} \qquad (62)$$

$$= G_{oo} - \frac{K_1(1+K_2G_{22})|G_{o1}|^2 + K_2(1+K_1G_{11})|G_{o2}|^2 - 2K_1K_2\mathrm{Re}(G_{o1}G_{12}G_{o2}^*)}{(1+K_1G_{11})(1+K_2G_{22}) - K_1K_2|G_{12}|^2} \qquad (63)$$

where Re( ) [illegible]

As before the [illegible] represents the effect of the interference. If, as before,

$$\underline{G}_o(n) = \underline{1}$$

then $G_{oo} = G_{11} = G_{22} = M$

$$|G_{os}|^2 = M + 2\sum_{i=1}^{M-1}\sum_{k=i+1}^{M} \cos\omega_n(\tau_k^{(s)} - \tau_i^{(s)}) = M + 2\sum_{i=1}^{M}\sum_{k=i+1}^{M} \cos\omega_n(\tau_i^{(s)} - \tau_k^{(s)})$$

$$|G_{12}|^2 = M + 2\sum_{i=1}^{M-1}\sum_{k=i+1}^{M} \cos\omega_n(\tau_i^{(1)} - \tau_i^{(2)} - \tau_k^{(1)} + \tau_k^{(2)})$$

and $$\mathrm{Re}(G_{o1}G_{12}G_{o2}^*) = \sum_{i=1}^{M}\sum_{k=1}^{M}\sum_{\ell=1}^{M} \cos\omega_n(\tau_i^{(1)} - \tau_k^{(1)} + \tau_k^{(2)} - \tau_\ell^{(2)}) \qquad (64)$$

If the two interference sources are widely separated in angle from each other and from the target, then $\tau_k^{(1)}$ and $\tau_k^{(2)}$ differ substantially for all k and are not close [illegible] the coefficient of K_1K_2 in both the numerator and denominator [illegible] increase with M while the other terms increase with M^2. Hence, for large M these coefficients become negligible with the result.

$$F(n) = [illegible] \qquad (65)$$

$$[illegible] \qquad (65a)$$

where the second approximation involves neglect of the oscillating [illegible] $|G_{o1}|^2$ and $|G_{o2}|^2$. Thus the effect of interferences is seen to be [illegible] under these conditions. For small interference-to-ambient-noise ratio, where K_1M and K_2M are very much less than unity, F(n) is reduced roughly by [illegible], while for very large interference-to-ambient-noise ratio, [illegible] no greater than 2. Thus for small interference-to-noise ratio the [illegible] index d decreases roughly with the first power of interference-to-[illegible] ratio (see Eq. 44), but the maximum effect is no greater than the loss of two hydrophones.

Suppose next that the two interference sources are sufficiently close together so that for all frequencies of interest, and for all i

$$\omega_n \left| \tau_i^{(1)} - \tau_i^{(2)} \right| \approx 0 \tag{66}$$

then $|G_{o2}|^2 \approx |G_{o1}|^2$

$$|G_{12}|^2 \approx M^2$$

and $\mathrm{Re}(G_{o1}G_{12}G^*_{o2}) \approx M|G_{o1}|^2$ (67)

$$\text{Then} \quad F(n) \approx G_{oo} - \frac{K_1(1+K_2M)|G_{o1}|^2 + K_2(1+K_1M)|G_{o1}|^2 - 2MK_1K_2|G_{o1}|^2}{1+(K_1+K_2)M}$$

$$= G_{oo} - \frac{(K_1+K_2)|G_{o1}|^2}{1+(K_1+K_2)M} \tag{68}$$

Thus the result converges to the case of a single plane-wave interference of strength $K_1 + K_2$ in this case.

For a linear array of M elements spaced d ft. apart we can set

$$\left| \tau_i^{(1)} - \tau_i^{(2)} \right| = \left| M/2 - i \right| \frac{d}{c} \left| \sin [illegible]_1 - \sin [illegible]_2 \right|$$

where the delay at the [illegible]

For small $\theta_2 - \theta_1$,

$$|\tau_i^{(1)} - \tau_i^{(2)}| = 2\,|\,M/_2 - i\,|\,\frac{d}{c}\sin\left|\frac{\theta_1-\theta_2}{2}\right|\cos\left(\frac{\theta_1+\theta_2}{2}\right)$$

$$\approx |\,M/_2 - i\,|\,\frac{d}{c}|\,|\theta_1 - \theta_2|\cos\theta_m \tag{70}$$

where $\theta_m = \frac{\theta_1 + \theta_2}{2}$ is the angle half-way between the two interferences. Then, since $\omega_{n_{max}}$ is $2\pi W$

$$\omega_n\left|\tau_i^{(1)} - \tau_i^{(2)}\right|_{max} = \pi WM\frac{d}{c}\,|\,\theta_1 - \theta_2|\cos\theta_m \tag{71}$$

and the two interferences are close enough together so that Eq.(68) holds if

$$\pi WM\frac{d}{c}\,|\,\theta_1 - \theta_2|\cos\theta_m \ll 1$$

or
$$|\theta_1 - \theta_2| \ll \frac{1}{\pi WM\frac{d}{c}\cos\theta_m} \tag{72}$$

As an example let $W = 5000$ cps, $d = 2$ ft, $c = 5000$/ft and $M=10$, then if $|\theta_1 - \theta_2| \ll .016/\cos\theta_m$ radians the two interference points have the same effect as a single one with a higher power level and therefore the maximum detectability loss can be no greater than $2/_3$ of a hydrophone as shown by Eq.(61).

Note that for $\theta_m = \pi/_2$ the two interferences are located symmetrically to the end-fire axis of the array; therefore their effect is always that of a single interference. This, however, is due to the symmetry of the linear array and does not hold in other cases.

Note further that Eq.(72) is a rather conservative limit since neither the effect of integration over frequency or over hydrophone spacing has been considered. Depending on the exact form of the power spectrum these integrations should result in increasing the value of $|\theta_1 - \theta_2|$ by a factor of 4 or 5 over that given in Eq.(72)

If the interference-to-noise ratio is small enough so that $(K_1 + K_2)M \ll 1$, then Eq.(68) and Eq.(65a) are approximately the same; thus under this condition

the effect of two interferences on the detectability is proportional to the interference power, and independent of the spacing of the two interference sources from each other. Note, however, that approximating $|G_{o1}|^2$ and $|G_{o2}|^2$ by M still implies that the interference direction is substantially different from the target direction.

VII. More than Two Interferences

Extension of the results obtained so far to more interferences is difficult and involves further approximations. Consider first the large interference-to-noise ratio case, with all interferences widely separated. We assume as before that $\underline{Q}_o = \underline{I}$ and that therefore $G_{rr} = M$ for $r = 0,1,2,...R$. It can be seen from Eq.(64) that the off-diagonal elements of $\underline{G}$ are of order $\sqrt{M}$. It can be shown in general (see Appendix B) that for the purpose of approximate inversion an n-dimensional matrix whose diagonal elements are of order k relative to the off-diagonal elements can be approximated by a diagonal matrix if $k >> n$. Thus if $\sqrt{M} >> R$ the off-diagonal elements of the matrix $[\underline{I} + \underline{G}]$ can be neglected in forming the inverse, with the result that

$$F(n) \approx G_{oo} - \sum_{r=1}^{R} \frac{K_r |G_{ro}|^2}{1 + K_r G_{rr}} \tag{73}$$

$$= M - \sum_{r=1}^{R} \frac{K_r |G_{ro}|^2}{1 + K_r M} \tag{74}$$

$$= M - \sum_{r=1}^{R} \frac{K_r M}{1+K_r M} \left[1 + \frac{2}{M} \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \cos \omega_n (\tau_i^{(r)} - \tau_k^{(r)})\right] \tag{75}$$

where $\tau_i^{(r)}$ is the interference delay from the r^{th} interference source at the i^{th} hydrophone. Eq.(73) is [illegible] Eq.(6[illegible]) and indicates that for

widely spaced point interference sources the detectability loss is approximately equivalent to the loss of one hydrophone per interference source. The approximation is good only for $R \ll \sqrt{M}$.

VIII. Effect of Distributed Interference Source

A distributed interference source can be represented by a large number of closely spaced point sources. Suppose that the interference source has a spectral density $I(n)$ and that the interference power is uniform for angles inside the interval $\theta_1 \leq \theta \leq \theta_2$ and zero outside. Then the interference can be represented by R points of spectral density $I(n)/R$ equally spaced in the interval, where R is a large number. Initially it will be assumed that the interference-to-ambient noise ratio is small. Although the result obtained under this assumption is somewhat academic (since the interference effect is very small in any case) it is possible to obtain an analytic result which is probably applicable with some modifications to larger interference-to-ambient-noise ratios as well. Under this assumption, the elements of the matrix $\underline{G}$ are all very small, and it is approximately true that

$$\underline{I} + \underline{G} \approx \underline{I} \tag{76}$$

Then the matrix inversion is, of course, trivial. The precise conditions for Eq.(76) to be a good approximation may be deduced from [illegible]; a simple sufficient condition is that

$$\sum_{s=1}^{R} \sqrt{K_r K_s}\, |G_{rs}| \ll 1 \qquad \text{for all } r = 1, \ldots.. R \tag{77}$$

In the present discussion $K_r = K_I/R$ for all $r = 1, \ldots,$ where $K_I = I(n)/N_o(n)$ is the total interferece-to-ambient-noise ratio. A conservative upper bound on K_I such that Eqs.(76) and (77) are good approximations is obtained by letting $|G_{rs}| = M$ for all r,s (see Eq.([illegible])). Hence, if

$$K_I M \ll 1 \tag{78}$$

Eq.(76) is a good approximation, and under these conditions Eq.(47) becomes

$$F(n) = G_{oo} - \underline{g}^T \underline{g}^*$$

$$= G_{oo} - \frac{K_I}{R} \sum_{r=1}^{R} |G_{or}|^2 \qquad (79)$$

and by use of Eq.(64) this becomes:

$$F(n) = G_{oo} - \frac{K_I}{R} \sum_{r=1}^{R} \{ M + 2 \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \cos \omega_n (\tau_i^{(r)} - \tau_k^{(r)})\} \qquad (80)$$

$$= G_{oo} - K_I M - \frac{2K_I}{R} \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \sum_{r=1}^{R} \cos \omega_n (\tau_i^{(r)} - \tau_k^{(r)})$$

We assume now that the azimuth angle subtended by the interference is small enough so that the $\tau_i^{(r)}$ do not differ very greatly as r goes from 1 to R. Then, it is possible to expand $\tau_i^{(r)}$ in a Taylor series in r as follows:

$$\tau_i^{(r)} = \tau_i^{(m)} + (r - m) \Delta \tau_i \qquad (81)$$

where $m = \frac{R}{2}$ is used as the point about which the expansion is performed; $\tau_i^{(m)}$ is effectively the mean delay of the interference wavefront.

As R is allowed to go to infinity the summation in r can be converted into an integral and evaluated as follows:

$$\sum_{r=1}^{R} \cos \omega_n (\tau_i^{(r)} - \tau_k^{(r)})$$

$$\approx \sum_{r=1}^{R} \cos \omega_n [\tau_i^{(m)} - \tau_k^{(m)} + (r-m)(\Delta\tau_i - \Delta\tau_k)]$$

$$\rightarrow \int_0^R \cos \omega_n [\tau_i^{(m)} - \tau_k^{(m)} + (r-m)(\Delta\tau_i - \Delta\tau_k)] dr$$

$$= R \frac{\sin \frac{R\omega_n}{2} (\Delta\tau_i - \Delta\tau_k)}{\frac{R\omega_n}{2} (\Delta\tau_i - \Delta\tau_k)} \cos \omega_n (\tau_i^{(m)} - \tau_k^{(m)}) \qquad (82)$$

Hence Eq.(80) becomes, after some reduction

$$F(n) = G_{oo} - K_I \left[M + 2 \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \frac{\sin \frac{\omega_n R}{2} (\Delta\tau_i - \Delta\tau_k)}{\frac{\omega_n R}{2} (\Delta\tau_i - \Delta\tau_k)} \cos\omega_n (\tau_i^{(m)} - \tau_k^{(m)}) \right] \tag{83}$$

$$= G_{oo} - K_I \left[M + 2 \sum_{i=1}^{M-1} \sum_{k=i+1}^{M} \frac{\sin \omega_n (\tau_i^{(m)} - \tau_k^{(m)})}{\omega_n (\tau_i^{(m)} - \tau_k^{(m)})} \cos \omega_n (\tau_i^{(m)} - \tau_k^{(m)}) \right] \tag{84}$$

where, in going from Eq(83) to (84) we have used the fact that

$$\frac{R}{2} (\Delta\tau_i - \Delta\tau_k) = m(\Delta\tau_i - \Delta\tau_k) = \tau_i^{(m)} - \tau_k^{(m)} \tag{85}$$

As before, the term representing the loss of detectability is the bracketed term in Eq.(84). Except for the $\frac{\sin x}{x}$ term the form of the double summation is the same as that which would be obtained for a single interference, (Eq. 54) with a mean delay $\tau_i^{(m)}$ at the i^{th} hydrophone, under the condition $K_I M \ll 1$. In fact, the argument that the summation of oscillating terms tends to become negligible applies here with ever greater force, because of the $\frac{\sin x}{x}$ term. One can conclude, therefore that for interference-to-ambient-noise ratio small enough to satisfy Eq.(78), and if the angle subtended by the interference is relatively small, a distributed interference source affects the performance in essentially the same way as a single point interference.

In order to obtain an estimate of the magnitude of azimuth angle that can be considered "small", consider a linear array with M hydrophones spaced d feet apart. For such an array

$$\tau_i^{(r)} = i \frac{d}{c} \sin \theta_r \tag{86}$$

where θ_r is the azimuth angle of the r^{th} interference point. Assume that the interference power is uniform over the range $\theta_1 \leq \theta \leq \theta_2$ and is zero outside this

range. The center of the interference is at the angle

$$\theta_m = \frac{1}{2}(\theta_1 + \theta_2) \tag{87}$$

Then, by analogy with Eq.(81) we expand $\sin\theta_r$ about θ_m; i.e.

$$\sin\theta_r \approx \sin\theta_m + (\theta_r - \theta_m)\cos\theta_m \tag{88}$$

All the other steps leading to Eq.(84) can then be performed in exactly the same way, with the summation over k replaced by an integration over θ_r. The final result can be put into the form

$$F(n) = G_{co} - K_I \left| M + 2\sum_{k=1}^{M}(M-k)\cos\left(\frac{k\omega_n d}{c}\sin\theta_m\right) \frac{\sin\left(\frac{k\omega_n d}{c}\cos\theta_m\left(\frac{\theta_2-\theta_1}{2}\right)\right)}{\frac{k\omega_n d}{c}\cos\theta_m\left(\frac{\theta_2-\theta_1}{2}\right)} \right| \tag{89}$$

Except for the $\frac{\sin x}{x}$ term in the summation, this is again the expression that we would have obtained for a single point interference located at the angle θ_m. It is clear that the accuracy of this expression depends on the accuracy of Eq.(88), which, in turn is a fairly good approximation for $\theta_2 - \theta_1$ less than about one radian. Thus we conclude that an interference source spread over no more than one radian affects the detectability essentially like a single point interference, provided that $K_I M << 1$.

Since the effect of interference for $K_I M << 1$ is very small, the result just obtained is not very interesting and it would be desirable to extend it somehow to the case of $K_I M >> 1$. Unfortunately this is quite difficult; in fact, the only simple result that has been obtained is an extension of Eq.(68) to more than 2 interference sources. If, as in the case of two interferences, it is assumed that the interference points are close enough together so that for all frequencies of interest

[illegible]

then $\quad |G_{or}|^2 \approx |G_{o1}|^2 \quad$ for all $r=1,\ldots,R$ (91)

and $\quad G_{rs}(n) \approx M \quad$ for all r,s,n. (92)

Then the matrix $\underline{I} + \underline{G}$ becomes

$$\underline{I} + \underline{G} \approx \begin{bmatrix} 1 + K_1 M & \sqrt{K_1 K_2} M & \cdots\cdots & \sqrt{K_1 K_R} M \\ \sqrt{K_2 K_1}\, M & 1 + K_2 M & \cdots\cdots & \sqrt{K_2 K_R} M \\ \vdots & & & \\ \sqrt{K_R K_1} M & \cdots\cdots & & 1 + K_R M \end{bmatrix} \tag{93}$$

$$= \underline{I} + \begin{bmatrix} \sqrt{K_1 M} \\ \sqrt{K_2 M} \\ \vdots \\ \sqrt{K_R M} \end{bmatrix} \begin{bmatrix} \sqrt{K_1 M} & \sqrt{K_2 M} & \ldots & \sqrt{K_R M} \end{bmatrix} \tag{94}$$

This is easily inverted by use of Eq.(17)

$$\left[\underline{I} + \underline{G}\right]^{-1} \approx \underline{I} - \frac{\underline{G}}{1 + M \sum_{r=1}^{R} K_r} \tag{95}$$

Also the vector $\underline{g}$ defined in Eq.(46) becomes:

$$\underline{g}^T = G_{o1} \left[\sqrt{K_1}\ \sqrt{K_2}\ \ldots\ \sqrt{K_R}\right]^T \tag{96}$$

and therefore $F(n)$ of Eq.(47) becomes

$$F(n) = M - \frac{|G_{o1}|^2 \sum_{r=1}^{R} K_r}{1 + M \sum_{r=1}^{R} K_r} \tag{97}$$

Obviously, if all the K_r are equal to K_I/R, Eq.(97) becomes

$$F(n) = M - \frac{|G_{o1}|^2 K_I}{1 + MK_I} \tag{98}$$

Thus the detectability loss is again equivalent to that of a single interference source of strength K_I, as one could have expected. It is clear from the result for the case of two point sources that for a linear array having M equally spaced hydrophones the maximum value of [illegible]$_1$ for which Eqs. (97) or (98) hold is given by Eq. (72).

IX. Computational Results

Since it has not been possible to obtain meaningful analytic results for cases in which the approximations made in the above work are not applicable, F(n) has been evaluated on a digital computer for a number of different array and interference patterns, and for specific frequencies. The results of some of these computations are presented in Figures 1 through 6. In all computations it is assumed that the array is steered on target at an angle $\theta = 0$ and that an interference exists at some angle θ_1. The curves are then plots of F(n) as θ_1 is varied. Thus, if θ_1 is near 0 the interference is near the target in azimuth, and F(n) is small. Also, the assumption that correlation of ambient noise waveforms between different hydrophones is zero has not been used; instead the exact form of the $\underset{\sim}{Q}_0(n)$ matrix as given by Bryn |3| was used. As a result $F(n) \neq M$ in the absence of interference as would be inferred from equations such as (52), (65), (75) or (98). In fact, $F(n) < M$ in all cases; however, this is a coincidence; it is possible for $F(n) > M$ as is shown by Bryn |3|. In all cases the interference-to-ambient-noise ratio is large.

Figure 1 shows the effect of a single point interference with a small circular array. It shows that if the interference directions differs by more than about 40° from the target direction the effect on F(n) is essentially negligible. It must be borne in mind, however, [illegible] demonstrated for a single frequency [illegible] different at other frequencies, and

the integrated effect of all frequencies therefore, has the effect of the loss of one hydrophone as is predicted by the analysis of Section V.

Figures (2) and (3) are similar to Figure 1 except that the interference consists respectively of two and of four points, separated by .1 radian. Since the interference covers a larger azimuth segment, the effect on F(n) covers a larger angle; however, it is still true that for interference sources at angles for removed from zero the effect on F(n) is small. A similar result is shown in Fig. (4) which shows the effect of two interferences separated by a large angle (90°). The last two figures show the effect of a strong interference (K_I=200) distributed over a relatively large angle (17°). Again the effect at angles far removed from the target angle is small, but, as is shown in Fig.(6), the relative effect is quite different at different frequencies, as has already been pointed out.

The computations leading to the results shown in Fig. 1 through 6 were quite time consumming, with computing times on the order of several minutes on the IBM 7094 for the cases with large numbers of interference points. For this reason no attempt was made to compute the complete detection index, since this would have required summation of F(n) over a large number of frequencies. The computer results therefore still don't conclusively answer the question of how serious the effect of large distributed interferences is. The indications are, however, that the results of Section VIII are valid under considerably wider conditions than those assumed there to produce analytical approximations. In fact, it appears that loss of detectability for rather widely distributed interference is equivalent to an [illegible]

X. Conclusions

The major difficulty in obtaining general estimates of the effect of directional noise on the detectibility in an array processor is that the mathematical manipulations required to obtain these results are quite complex. Results have therefore been obtained only in a restricted number of simple cases.

The general tenor of these results is that if the anisotropic-to-isotropic-noise ratio is small the effect of a number of local noise sources is additive; that is the loss of detectability resulting from two noise sources of equal strength is twice that resulting from a single source. For large anisotropic-to-ambient-noise ratio the effect depends on whether the directional noise sources are close together or not. For a single point source it has been shown previously and corroborated here, that the loss in detectability is approximately equivalent to the loss of one hydrophone from the array. If there are R noise sources, widely separated from each other and from the target direction the loss is approximately equivalent to the loss of R hydrophones, provided that R << M, where M is the number of hydrophones.

Point noise sources that are close together affect the system like a single distributed noise source, and the indications are that if such an anisotropy is spread over a relatively small angle its effect is essentially that of a single point noise. Unfortunately this has not been conclusively demonstrated, even by use of a digital computer, and only a rather conservative estimate of azimuth angle that can be considered to be "small" has been obtained.

Appendix A Derivation of the Detection Index

The detection index d is given by

$$d = \frac{\langle u \rangle_{S+N} - \langle u \rangle_N}{\sigma_N(u)} \qquad \text{(A-1)}$$

$$\text{where } u = \sum_{n=1}^{WT} \left| \underline{H}^T(n)\, \underline{x}(n) \right|^2 \qquad \text{(A-2)}$$

$$\text{and where } \underline{H}(n) = \sqrt{K(n)}\, \underline{Q}^{-1}(n)\, \underline{V}^*(n) \cdots K(n)\, \underline{V}^{*T}(n)\, \underline{Q}^{T-1}(n) \qquad \text{(A-3)}$$

$$\text{with} \quad K(n) = \frac{S(n)/N^2(n)}{1 + S(n)G_o(n)/N(n)} \qquad \text{(A-4)}$$

$$\text{Then } \langle u \rangle_N = \sum_{n=1}^{WT} \underline{H}^T(n) \langle \underline{X}(n)\, \underline{X}^{*T}(n) \rangle_N\, \underline{H}^*(n) \qquad \text{(A-5)}$$

$$\text{But from Eq. (7)} \quad \langle \underline{X}(n)\, \underline{X}^{*T}(n) \rangle_N = N(n)\, \underline{Q}^*(n) = N(n)\, Q^T(n) \qquad \text{(A-6)}$$

$$\text{Therefore } \langle u \rangle_N = \sum_{n=1}^{WT} N(n)\, \underline{H}^T(n)\, \underline{Q}^T(n)\, \underline{H}^*(n) = \sum_{n=1}^{WT} N(n)\, K(n)\, \underline{V}^{*T}(n)\, \underline{Q}^{T-1}(n)\, \underline{V}(n)$$

$$= \sum_{n=1}^{WT} K(n)\, N(n)\, G_o(n) \qquad \text{(A-7)}$$

Similarly, and using the fact that

$$\langle \underline{X}(n)\, X^{*T}(n) \rangle_{S+N} = N(n)\, \underline{Q}^T(n) + S(n)\, \underline{\Gamma}^T(n) \qquad \text{(A-8)}$$

$$\langle u \rangle_{S+N} = \sum_{n=1}^{WT} K(n) \left[S(n)\, G_o^2(n) + N(n)\, G_o(n) \right] \qquad \text{(A-9)}$$

Hence

$$\langle u \rangle_{S+N} - \langle u \rangle_N = \sum_{n=1}^{WT} K(n)\, S(n)\, G_o^2(n) \qquad \text{(A-10)}$$

To find $\sigma_N(u)$ it is necessary to find $\langle u^2 \rangle_N$, given by

$$\langle u^2 \rangle_N = \sum_{n=1}^{WT} \sum_{m=1}^{WT} \left\langle \left| \underline{H}^T(n)\, \underline{X}(n) \right|^2 \left| \underline{H}^T(m)\, \underline{X}(m) \right|^2 \right\rangle_N \qquad \text{(A-11)}$$

Let $w(n) = \underline{H}^T(n)\, \underline{X}(n)$, then $w(n)$ is a Gaussian random variable since $\underline{X}(n)$ is. In terms of $w(n)$

$$\langle u^2 \rangle_N = \sum_{n=1}^{WT} \sum_{m=1}^{WT} \left\langle w(n)\, w^*(n)\, w(m)\, w^*(m) \right\rangle_N$$

$$= \sum_{n=1}^{WT} \sum_{m=1}^{WT} \Big[\langle w(n)\, w^*(n) \rangle_N \langle w(m)\, w^*(m) \rangle_N + \langle w(n) w(m) \rangle_N \langle w^*(n) w^*(m) \rangle_N$$

$$+ \langle w(n)\, w^*(m) \rangle_N \langle w^*(n)\, w(m) \rangle_N \Big] \qquad \text{(A-12)}$$

This expansion is permissible because $w(n)$ is Gaussian [4]. The first term in the expansion is simply the square of the mean $\langle u \rangle_N^2$ the second term vanishes [5], and in the last term all terms for which $n \neq m$ vanish because components at different frequencies are assumed to be independent. Hence

$$\sigma_N^2(u) = \langle u^2 \rangle_N - \langle u \rangle_N^2 = \sum_{n=1}^{WT} \left\langle |w(n)|^2 \right\rangle_N^2 = \sum_{n=1}^{WT} \left[\underline{H}^T(n) \langle \underline{X}(n) \underline{X}^{*T}(n) \rangle_N \underline{H}^*(n) \right]^2$$

$$= \sum_{n=1}^{WT} K^2(n)\, N^2(n)\, \sigma_o^2(n) \qquad \text{(A-13)}$$

as in Eq.(A-7).

Thus, finally

$$d = \frac{\sum_{n=1}^{WT} K(n)\, S(n)\, \sigma_o^2(n)}{\sqrt{\sum_{n=1}^{WT} K^2(n)\, N^2(n)\, \sigma_o^2(n)}} \qquad \text{(A-14)}$$

For small signal-to-noise ratio, such that $S(n)\sigma_o(n)/N(n) << 1$

$$K(n) \approx S(n) / N^2(n) \qquad \text{(A-15)}$$

Then

$$d \approx \frac{\sum_{n=1}^{WT} \frac{S^2(n)\, \sigma_o^2(n)}{N^2(n)}}{\sqrt{\sum_{n=1}^{WT} \frac{[illegible]}{[illegible]}}} [illegible] \qquad \text{(A-16)}$$

Appendix B

Approximate Inversion of a Matrix whose Diagonal Terms are Large Relative to the Off - Diagonal Terms.

Let the n×n nonsingular matrix $\underline{A}$ be given by

$$\underline{A} = \underline{D} + \underline{B} \qquad \text{(B-1)}$$

where $\underline{D}$ is diagonal and $\underline{B}$ is a matrix with zero diagonal elements. It is assumed that all the non-diagonal elements of $\underline{B}$ are of about the same order of magnitude, and that the elements of $\underline{D}$ are of about K times that magnitude, with $K >> 1$.

The inverse of $\underline{A}$ is given by

$$\underline{A}^{-1} = (\underline{D} + \underline{B})^{-1} = \underline{D}^{-1}(\underline{I} + \underline{BD}^{-1})^{-1} = \underline{D}^{-1}\left[\underline{I} - \underline{BD}^{-1} + (\underline{BD}^{-1})^{2} \ldots\right] \qquad \text{(B-2)}$$

Since the elements of $\underline{D}$ are of order K relative to $\underline{B}$ the elements of $\underline{B}\,\underline{D}^{-1}$ are of order $^1/K$ relative to unity.

It can be shown [8] that a sufficient condition for convergence of Eq.(B-2) is

$$\sum_{j=1}^{n} \left| b_{ij} \right| < 1 \qquad i = 1.....n \qquad \text{(B-3)}$$

where b_{ij} are the elements of BD^{-1}. Assuming all of these elements to be of about the same order of magnitude, condition (B-3) can be expressed in the approximate form

$$nb_o < 1 \qquad \text{(B-4)}$$

where b_o is a representative element of BD^{-1}. This element is of order $^1/K$; therefore convergence requires

$$^n/K < 1 \qquad \text{(B-5)}$$

The convergence will clearly be more rapid if this inequality is sharper; hence one can approximately neglect the matrix $\underline{B}$ in the inversion of $\underline{A}$ if $^n/K << 1$.

L. t of References:

1. Schultheiss, Peter M. "Passive Detection of a Sonar Target in a Background of Ambient Noise and Interference from a Second Target", Progress Report No. 17, General Dynamics/Electric Boat Research, September, 1964. (This report is also in "Processing of Data from Sonar Systems, Vol. III, by Kanefsky, Levesque, Schultheiss, and Tuteur, General Dynamics/Electric Boat Research and Development Department, Groton, Connecticut, Report No. U 417-65-033 (August 23, 1965).

2. Edelblute, David J., Joanne M. Fisk, and Gerald L. Kinnison, "Criteria for Optimum-Signal-Detection Theory for Arrays." JASA 41 No. 1 (January 1967) pp. 199 to 205.

3. Bryn, F. "Optimum Signal Processing of Three-Dimensional Arrays Operating on Gaussian Signals and Noise". JASA 34 No.3 (March 1962) page 289.

4. Edelblute, et al., op cit., Eq.(41)

5. Knapp, Charles H. "Optimum Linear Filtering for Multi-element Arrays" General Dynamics/Electric Boat Division Technical Report U 417-66-031 (November 1966).

6. McDonald, R.A., P.M. Schultheiss, F.B. Tuteur, T. Usher, Jr. "Processing of Data from Sonar Systems," General Dynamics/Electric Boat Research and Development Department, Groton, Connecticut., Report No. U 417-63-045 (September 1, 1963.

7. Freeman, B. "Principles and Techniques of Applied Mathematics," John Wiley and Sons, N.Y., 1957, page 34.

8. Faddeeva, U.N. "Computational Methods in Linear Algebra" Dover Publications, New York, 1959.

9. Helstrom, C. "Statistical Theory of Signal Detection", Pergamon Press, 1960, page 53.

METHODS OF STOCHASTIC APPROXIMATION APPLIED TO THE ANALYSIS OF ADAPTIVE TAPPED DELAY LINE FILTERS

by

John H. Chang
Franz B. Tuteu

Progress Report No. 34

General Dynamic/Electric Boat Research

November 1967

DEPARTMENT OF ENGINEERING
AND APPLIED SCIENCE

YALE UNIVERSITY

ABSTRACT

In this report some studies have been carried out for a class of adaptive filters consisting of tapped delay lines and adjustable gains. The method of stochastic approximation and mean square error criterion are employed to adjust the gains automatically. It is shown that it is not necessary that the desired signal generally used to obtain the error function is available. Either signal or noise correlation functions will suffice to generate the error gradient. Problems basic to all adaptive processes such as the conditions for convergence, rate of convergence, effect of misadjustment, effect of time-varying parameters, and the relationship between mean square error and the number of delay elements are answered with explicit expressions.

Contents

I. INTRODUCTION

1.1 Optimum Linear Filters

The problem of designing a device to eliminate noise or to predict the future behavior of an incoming signal has been considered by Norbert Wiener[1] more than twenty years ago. This kind of device has been termed as "filter" in general. Consider a linear filter shown in Figure 1, where the input $x(t)$ is a combination of the useful signal $s(t)$ and noise $n(t)$. Assuming that $n(t)$ is additive to and statistically independent of $s(t)$, we have

$$x(t) = s(t) + n(t) \tag{1.1}$$

and

$$\overline{s(t)\, n(t)} = 0 \quad \text{if} \quad \overline{s(t)} = 0, \ \overline{n(t)} = 0 \tag{1.2}$$

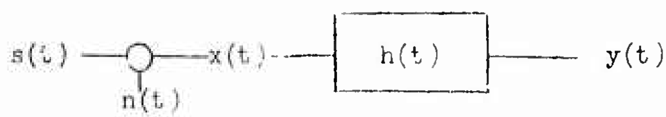


Figure 1. A linear filter

The output $y(t)$ of the filter is to approximate a desired function $d(t)$ which is related to the signal $s(t)$. The performance criterion to be minimized is the mean square error

$$\overline{e^2(t)} = \overline{\left(d(t) - y(t)\right)^2} \tag{1.3}$$

This is the classical problem of Wiener, the analytic solution (for the impulse response of a realizable filter) is known to be the solution of the Wiener-Hopf integral equation[2]

$$R_{xd}(\tau) = \int_0^{\infty} h(t)\, R_{xx}(\tau - t)\, dt \tag{1.4}$$

with the solution*

$$H_o(w) = \frac{\Phi_{xd}(\omega)}{\Phi_{xx}(\omega)} \tag{1.5}$$

and the minimum mean square error

$$\overline{e^2_{min}} = \int_{-\infty}^{+\infty} \left[\Phi_{dd}(\omega) - |H_o(\omega)|^2 \Phi_x(\omega) \right] d\omega$$

$$= \overline{d^2(t)} - \overline{y_o^2(t)} \tag{1.6}$$

In the above equations $R_{xy}(\tau)$ and $\phi_{xy}(w)$ are the cross correlation function and spectral density functions between $x(t)$ and $y(t)$. $y_o(t)$ is the output of the optimum filter. The results are valid for stationary signals.

Kalman and Bucy[3] have presented a new method to design optimal linear filters for nonstationary signals. They considered the model for the signal process as

$$\frac{d\underline{x}}{dt} = \underline{F}\underline{x} + \underline{G}u \tag{1.7}$$

where $u(t)$ is white noise, i.e.,

$$\overline{u(t)\, u(\tau)} = N\delta(t-\tau) \tag{1.8}$$

The observed signal is assumed as

$$\underline{z} = \underline{H}\underline{x} + v \tag{1.9}$$

where $\underline{H} = (1, 0)$ and v is white noise of spectral density ϕ.

*This type of filters may not be physically realizable. Further discussions can be found in Wiener[1]

The optimum filtering problem consists in determining

$$\hat{x}(t) = E\left\{ \underline{x}(t) \mid z(\tau) \right\} \tag{1.10}$$

where $\hat{x}(t)$, the conditional expectation of $x(t)$ given the observation $\underline{z}(\tau)$ in the interval $(0,t)$, is the minimum variance unbiased estimator of $x(t)$.

The optimum system is described succinctly by the following four equations[4]:

$$\begin{aligned}
&\text{with} \quad \frac{d\hat{\underline{x}}}{dt} = F\hat{\underline{x}} + PH^T\phi^{-1}(\underline{z} - H\hat{\underline{x}}) \\
&\qquad\quad \hat{\underline{x}}(o) = 0 \\
&\qquad\quad \frac{dP}{dt} = FP + PF^T - PH^T\phi^{-1}HP + \begin{pmatrix} 0 & 0 \\ 0 & N \end{pmatrix} \\
&\text{and} \quad\;\; P(0) = E\left\{ \underline{x}(o)\,\underline{x}^T(o) \right\}
\end{aligned} \tag{1.11}$$

For time varying systems, Kalman-Bucy filters can provide much better performance than Wiener filters. From (1.5) and (1.11) it is seen that the statistical properties of both signal and noise should be known in order to design either type of the two filters.

Thus, if the a priori information is not known completely, optimal performance of the filters cannot be expected. In an attempt to recover some of the missed a priori information by evaluation of the actual performance of the operating system, the concept of adaptation has been developed and accepted as one of the possible solutions to many such problems.

1.2 Adaptive Systems and Techniques

A system is adaptive if there is a decision-making box in the

feedback loop where the observable output is compared with the desired output so as to adjust the system for better performance[5, 6]. In other words, an adaptive system is one which is provided with a means of continuously monitoring its own performance according to a given performance index, and also a means of adjusting its own parameters by closed loop action so as to optimize its own operation. Relaxation method and method of steepest descent (or ascent) are two of the most commonly-used adaptive techniques.

The relaxation method involves making a change in the value of only one of the controller parameters and then re-evaluating the performance measure. If the performance has been improved, a second change in the same direction is made; otherwise, the first change is retracted and a change in the opposite direction is made. This process is continued until no further improvement in the performance measure can be accomplished by adjusting that particular parameter; whereupon the same process is repeated for each of the remaining controller parameters. After several iterations through the entire procedure, the controller parameters tend toward that set of values which yields the optimum performance measure.

The methods of steepest descent (or ascent), referred as gradient techniques are operated in a manner similar to the relaxation method, with the notable exception that all parameters are adjusted simultaneously rather than sequentially. This is done by measuring the partial derivative of the performance measure with respect to each of the controller parameters and then adjusting all the parameters in such a way that the net effect is the largest possible improvement in the performance measure.

A number of techniques have been developed for determining the partial derivatives.

The most straightforward method is to perturb each of the parameters sequentially and measure the derivatives directly. This procedure, however, offers little advantage over the relaxation method. A second technique is to perturb the parameters simultaneously in such a manner that the effect of the perturbation of each parameter on the performance measure will be distinguishable from the effects of the perturbations of all the other parameters. Ways in which this may be done include perturbation by independent random noise, distinguishing the individual effects by correlation detection[7]; or perturbation by frequency-separated sinusoide, distinguishing the effects by narrow-band detection[8]. Gradient techniques can be considered as the special case of the more general method of stochastic approximation, by which either deterministic or random problems can be solved with ease.

1.3 Adaptive Filters and State of the Art

Adaptive filters have been investigated by a number of researchers[4, 9, 10]. Their methods differ chiefly in the ways of implementation, but all are designed with the same purpose in mind - to extremize the performance index by gradually adjusting the system parameters. One of the simplest implementations in this area is the use of tapped delay lines which can be constructed easily with shift registers in digital computers. Weaver[10] considered tapped delay line filters, but his adaptive scheme was rather ineffective and required the solution of many simultaneous equations. Narendra and the author [11] used delay lines to identify the characteristics of some unknown nonlinear dynamic systems. Although the

feasibility of the above-mentioned methods was indicated by computer simulations, some figures of merit to judge these schemes, such as the rate of convergence of the parameters to the optimum, remain untackled. Widrow[12, 13] has attempted to solve these problems by defining some adaptive constants and misadjustment formulas. His system consists of delay lines and adjustable gains, and has been acclaimed to perform nearly as the Kalman-Bucy filter when perfect a priori information is available. Under circumstances in which the a priori information is not perfectly known, it is quite possible that the performance of such an adaptive filter could exceed that of either a Wiener or a Kalman-Bucy filter. However, Widrow's system requires the availability of a desired signal to generate the real time error function. Convergence proof of his LMS adaptation algorithm (least-mean-square-error algorithm) was not given, and the development of the rate of adaptation was not mathematically rigorous. Moreover, how to make a time-varying system adaptive has not been considered.

1.4 Outlines of the Report

A. Consider a random function $Q(\underline{x}|\underline{c})$ where $\underline{x} = \{x_1, x_2, \ldots, x_n\}$ is a vector of stationary random process with distribution $P(\underline{x})$. In an attempt to minimize the criterion

$$I(\underline{c}) = E_x \left(Q(\underline{x}|\underline{c}) \right) \tag{1.12}$$

it is natural to set the gradient of $I(\underline{c})$ to zero,

$$\nabla I(\underline{c}) = E_x \left\{ \nabla_c Q(\underline{x}|\underline{c}) \right\} = 0 \tag{1.13}$$

Since $P(\underline{x})$ is generally unknown, an algorithm derived from the method of stochastic approximation to obtain $\underline{c}^*$, the optimum value of $\underline{c}$, is

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \nabla_c Q(\underline{x}_j|\underline{c}_j) \tag{1.14}$$

Algorithm (1.14) converges with probability one

$$P\left\{ \lim_{j\to\infty} (\underline{c}_j - \underline{c}^*) = 0 \right\} = 1 \tag{1.15}$$

as well as in mean square

$$\lim_{j\to\infty} E\left\{ \|\underline{c}_j - \underline{c}^*\|^2 \right\} = 0 \tag{1.16}$$

under the following conditions

$$\text{A.}\quad \lim_{j\to\infty}\gamma_j = 0,\ \sum_{j=1}^{\infty}\gamma_j = \infty,\ \sum_{j=1}^{\infty}\gamma_j^2 < \infty,\ \gamma_j > 0 \tag{1.17}$$

$$\text{B.}\quad \inf_{\varepsilon < \|\underline{c}-\underline{c}^*\| < \frac{1}{\varepsilon}} E\left\{ (\underline{c} - \underline{c}^*)^T \nabla_c Q(\underline{x}|\underline{c}) \right\} > 0 \qquad \varepsilon > 0 \tag{1.18}$$

$$\text{C.}\quad E\left\{ \nabla_c^T Q(\underline{x}|\underline{c}) \nabla_c Q(\underline{x}|\underline{c}) \right\} \le d(\underline{c}^{*T}\underline{c}^* + \underline{c}^T\underline{c}) \tag{1.19}$$

for all $\underline{c}$ and $d > 0$

B. For the tapped delay line filter under study, the transfer function of the filter is represented by

$$H(\omega) = \sum_{k=0}^{N} c_k e^{-j\omega T_k} \tag{1.20}$$

and its impulse response is

$$h(t) = \sum_{k=0}^{N} c_k \delta(t - T_k) \tag{1.21}$$

If we attempt to minimize the mean square error

$$e^2 = E\left\{ \left[d(t) - z(t)\right]^2 \right\} = E\left\{ \left(d(t) - \sum_{k=0}^{N} c_k\, x(t-T_k)\right)^2 \right\} \tag{1.22}$$

the optimum values of $\underline{c}$ is obtained as

$$\underline{c}^* = \underline{R}_\eta^{-1}\, \underline{R}_{d\eta}$$

where (1.23)

$$\underline{R}_\eta = E\begin{pmatrix} \eta_0\eta_0 & \cdots & \eta_0\eta_N \\ \eta_N\eta_0 & & \eta_N\eta_N \end{pmatrix}, \quad \underline{R}_{d\eta} = E\begin{pmatrix} d\eta_0 \\ d\eta_N \end{pmatrix}$$

with $\eta_k(t) = x(T-T_k)$

it is seen that in (1.23) the values of $\underline{c}^*$ cannot be determined unless we have full knowledge about both signal and noise correlations.

Substituting (1.23) into (1.22) we can obtain the expression for the minimum mean square error as

$$\begin{aligned} \overline{e^2_{min}} &= \overline{d^2(t)} - \underline{R}_{d\eta}^T\, \underline{c}^* \\ &= \overline{d^2(t)} - \underline{R}_{d\eta}^T\, \underline{R}_\eta^{-1}\, \underline{R}_{d\eta} \\ &= \overline{d^2(t)} - \overline{z_0^2(t)} \end{aligned} \tag{1.24}$$

where $z_0(t)$ is the output of the optimum filter.

The mean square error at any time for arbitrary values of $\underline{c}$ is

$$\overline{e^2(t)} = \overline{e^2_{min}} + (\underline{c} - \underline{c}^*)^T\, \underline{R}_\eta\, (\underline{c} - \underline{c}^*) \tag{1.25}$$

and the effect of non-optimum setting is bounded by

$$\overline{e^2(t)} - \overline{e^2_{min}} \leq (N+1)^2 \max_{\text{all } i} \left| c_i - c_i^* \right| \max_{i,j} \left| \overline{\eta_i\eta_j} \right| \tag{1.26}$$

A relationship between the minimum mean square error and the number of delay elements is

$$\overline{e^2_{min}} = R_s(o) - \underline{R}_s^T \ \underline{R}_\eta^{-1} \ \underline{R}_s \tag{1.27}$$

The last term in the right-hand side of (1.27) is a functional of N, the number of delay elements and the correlation functions. For any known forms of $R_n(\tau)$ and $R_s(\tau)$, a plot of $\overline{e^2_{min}}$ versus N can be constructed. It is anticipated that the larger N is, the smaller $\overline{e^2_{min}}$ will be.

C. Under various practical situations we may not have perfect information about both $R_s(\tau)$ and $R_n(\tau)$. Techniques of adaptation can be employed to estimate the incomplete a priori information and to make the filter adaptive to changing operating conditions.

From (1.22) the error gradient without averaging operation is obtained as

$$\nabla_c \ Q(\underline{x}|\underline{c}) = -2 \ (\ d(t) - z(t) \ ) \underline{\eta}(t)$$

so that the desired adaptive algorithm is

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \ e_j \underline{\eta}_j \tag{1.28}$$

The above algorithm converges if 1. Q(c) is strictly convex, 2. $\frac{\partial^2 Q}{\partial \underline{c}^2}$ exists and is uniformly bounded, 3. s(t) and n(t) are uniformly bounded.

In (1.28) the desired signal d(t) is used to generate the error function. This assumption is not so practical when dealing with

detection problems. If only the noise correlation function is known, we can change Eq. (1.28) to

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \underline{\eta}_j (x_j - z_j) - 2\gamma_j \underline{R}_n \tag{1.29}$$

where $\underline{R}_n^T = \left(R_s(o), \ldots, R_s(T_n) \right)$

On the other hand, if only the signal correlation function is known, we have

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \underline{R}_s - 2\gamma_j z_j \underline{\eta}_j \tag{1.30}$$

where $\underline{R}_s^T = \left(R_s(o), \ldots, R_s(T_N) \right)$

D. The other problem that we have to consider is the rate of adaptation, or the rate of convergence. We want to estimate how fast the gains approach their optimum values. Defining

$$E\left\{(\eta)(\eta)^T\right\} = E\{\underline{H}\} = \underline{P}^{-1} \Lambda \underline{P} \tag{1.31}$$

$$\underline{W} = \underline{P} \underline{c}, \quad \eta' = \underline{P} \eta, \tag{1.32}$$

and $\gamma_j^p = \dfrac{1}{2(j+1)\lambda_p}$ for the p^{th} component of W,

we can express the component of $\underline{W}$ at any time during the adaptation period as a function of the initial choice and the optimum values of $\underline{W}$, i.e.,

$$w_j = \frac{1}{j+1} (w_1 - w^*) + w^* \tag{1.33}$$

Using (1.24) we obtain the difference between $\overline{e^2(t)}$ at any time and $\overline{e^2_{min}}$ as

$$\overline{e^2_j} - \overline{e^2_{min}} = \sum_{p=0}^{N} \lambda_p \left(\frac{w_1^p}{j} - \frac{w_1^p}{j}\right)^2 \tag{1.34}$$

$$= \frac{(N+1)\overline{e}^2{}_{min}}{j} \frac{1}{j^2} \left\{ (\underline{c}_1 - c^*)^T \underline{R}_\eta (\underline{c}_1 - \underline{c}^*) \right\}$$

$$\hat{t} = \beta \tilde{t} \tag{1.38}$$

then the method of two time scales can be used to modify the algorithms obtained previously. The mean square error is changed to

$$\overline{e^2(t)} \cong E\left\{\left[d(t) - \sum_{i=0}^{N} c_i \;{}_i(t)\right]^2\right\}$$

$$+ 2\beta T_{av} E\left\{\left(d(t) - \sum_{i=0}^{N} c_i \;{}_i(t)\right)\sum_{i=0}^{N} \eta_i(t) \frac{\partial c_i(\hat{t})}{\partial \hat{t}}\right. \tag{1.39}$$

where T_{av} is the average delay time. Algorithm (1.28) is modified to the form

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \underline{\eta}_j e_j + 2\gamma_j \beta T_a \underline{\eta}_j \underline{\eta}_j^T \underline{\delta} \tag{1.40}$$

where

$$\delta_i = \frac{\partial c_i(\hat{t})}{\partial \hat{t}}$$

The minimum mean square error for this case is

$$\overline{e^2_{min}} \leq \overline{e^2_{o\,min}} + \beta T_{av} \left| \underline{\delta}^T \right| \underline{R}_{d\eta} + 2\beta^2 T_{av}^2 \left| \underline{\delta}^T \underline{R}_{\eta} \underline{\delta} \right| \tag{1.41}$$

where $\overline{e^2_{o\,min}}$ is the minimum mean square error of the time-invariant filters.

1.5 Research Work in Progress

So far preliminary results have been obtained for a single input single output filter under very general situations. Practical examples for different types of signal, noise and ways of parameter variation are being worked out. In order to verify the results, digital computer simulation will be conducted.

where λ_p is the p^{th} component of $\underline{\Lambda}$ and w_i^p is the p^{th} component of $\underline{w}_1$.

E. So far w e have assumed that the whole system is time-invariant and the output signals are stationary at least in the wide sense. However, some times for one reason or another, system parameters or signal properties may change slowly. One instance of this situation is the fluctuation of power levels. Some schemes to adjust the gains under this case would be highly desirable. First consider the quasi-stationary case where only slow time variable is involved. If a time function $x(t)$ is delayed by an amount of T and multiplied by c such that

$$y(t) = cx(t-T)$$

or

$$y(t+T) = cx(t)$$

which can be written as

$$Ly(t) = cx(t) \tag{1.35}$$

L is a linear differentiation operator

$$L = e^{pT} = \sum_{i=0}^{\infty} \frac{T^i}{i!} p^i \quad \text{and} \quad p = \frac{d}{dt} \tag{1.36}$$

Let $x(t) = e^{\lambda t}$ and $y(t) = z(t,\lambda)e^{\lambda t}$, then $z(t,\lambda)$ can be obtained as $z_o(t,\lambda) = ce^{-\lambda t}$

if c is a constant, and

$$z(t,\lambda) = \frac{1}{e^{\lambda t}} \left(c - \sum_{i=1}^{n} \frac{T^i}{i!} \frac{d^i c(t)}{dt^i} \right) \tag{1.37}$$

if c is time-varying. The effect of time-varying parameter is observed.

During the training period the system is operated in real (fast) time $\tilde{t}$. If at the same time some parameters are changing slowly in slow time $\hat{t}$ such that

A very interesting and practical application of adaptive filters is the sonar detecting system consisting of many hydrophones steered or not steered on target. The output of each hydrophone passes through an adaptive filter, and the sum of the filter outputs is squared and averaged to indicate the presence or absence of a target. When the input signal to noise ratio is small and the levels of signal and noise are the same at all hydrophones, the whole system can be adjusted to form a likelihood ratio detector. Some features of this detector will be studied with emphasis on adaptive schemes, rate of convergence and detectability for stationary and time-varying processes. Performance analysis will be carried out with analytic and numerical examples.

II. METHODS OF STOCHASTIC APPROXIMATION

1. Historical Developments

The methods of stochastic approximation were originally developed by Robbins and Monro in 1951.[14] Their purpose was to find the root of a noisy function. The term "stochastic" refers to the random character of the experimental errors, while the term "approximation" refers to the continued use of past measurements to estimate the approximate position of the goal. Kiefer and Wolfowitz[15] adapted the idea of stochastic approximation to the problem of finding the maximum of a unimodal function obscured by noise. Blum[16] used the gradient method to extend the above techniques to multi-dimensional case. Later on Dvorestzky[17] greatly generalized and unified the whole theory and Kesten[18] derived some formulas to speed up the rate of convergence in terms of the number of changes in sign before a certain step.

2. Basic Considerations[19]

Stochastic approximation, much like ordinary successive approximation in the absence of experimental error, involves two basic considerations - first choosing a promising direction in which to search and selecting the distance to travel in that direction. Picking a search direction is no more difficult for stochastic than for deterministic approximations, for one simply behaves as if he believed the experimental results, ignoring entirely the possibility of error. This means of course that the experimenter will move away from his goal whenever he is misled by the vagaries of chance error. It will be seen that such temporary set-backs do not prevent

ultimate convergence if the step sizes are chosen properly.

In both stochastic and deterministic schemes, the corrections are made progressively small as the search proceeds so that the process will eventually converge. To make this convergence rapid, one would like to shrink the step size as speedily as possible. The main difference between stochastic and deterministic procedures is in fact the speed with which the steps can be shortened. When noise is totally absent one can reduce the steps very rapidly, but when there is danger of an occasional jump in the wrong direction, shortening the steps too rapidly could make it impossible to erase the long-run effects of a mistake. In the latter case the process would still converge, but to the wrong value.

3. The Methods[20]

Many problems in modern cybernetical systems design can be reduced to that of finding the extrema of functions of several variables

$$I = Q(c_1, c_2, \ldots, c_n) = Q(\underline{c}) \quad \text{where } \underline{c} = \{c_1, c_2, \ldots c_n\} \tag{2.1}$$

Denoting the optimal values of $\underline{c}$ by $\underline{c}^*$ and assuming that the extremum of interest to us is a minimum, we can obtain the solution of $\underline{c} = \underline{c}^*$ by setting the gradient of $Q(\underline{c})$ equal to zero; i.e.,

$$\nabla Q(\underline{c}) = 0 \tag{2.2}$$

$$\text{where} \quad \nabla Q(\underline{c}) = \left\{ \frac{\partial Q(\underline{c})}{\partial c_1}, \ldots, \frac{\partial Q(\underline{c})}{\partial c_n} \right\}$$

Generally a closed-form solution cannot be obtained for (2.2), so iteration methods are required, especially the gradient method.

The gradient method relates the coordinates of a given point with the coordinates of the preceding point and the gradient $\nabla Q(\underline{c})$. The algorithm for determining $\underline{c}^*$ can be written in the form

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \nabla Q(\underline{c}_j) \tag{2.3}$$

Here γ_j determines the pitch of the algorithm and generally depends on the index of the step and the function itself.

When $Q(\underline{c})$ is not given analytically or is not differentiable, the gradient $\nabla Q(\underline{c})$ can be approximately determined with the formula

$$\frac{Q_+(\underline{c}, a) - Q_-(\underline{c}, a)}{2a}$$

where

$$Q_{\pm}(\underline{c}, a) = \left\{ Q(\underline{c} \pm a\underline{e}_1), \ldots, Q(\underline{c} \pm ae_n) \right\} \tag{2.4}$$

and e_i denotes the base vectors

$$\underline{e}_1 = \left\{1, 0, \ldots, 0\right\}, \; e_n = \left\{0, 0, \ldots, 1\right\}. \tag{2.5}$$

The corresponding algorithm is then

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \left\{ \frac{Q_+(\underline{c}_j, a_j) - Q_-(\underline{c}_j, a_j)}{2a_j} \right\} \tag{2.6}$$

In the above we assumed that $Q(\underline{c})$ is a deterministic function. If we consider a random function $Q(\underline{x}|\underline{c})$, where $\underline{x} = \left\{ x_1, x_2, \ldots, x_n \right\}$ is a vector of stationary random processes with distribution $p(\underline{x})$, it is natural to attempt to find the extrema of the mathematical expectation:

$$I(\underline{c}) = \int_{\underline{X}} Q(\underline{x}|\underline{c})\, p(\underline{x})\, d\underline{x} = E_x \left\{ Q(\underline{x}|\underline{c}) \right\}. \tag{2.7}$$

The condition for determining the optimal value $\underline{c} = \underline{c}^*$ is of the form

$$\nabla I(\underline{c}) = E\left\{ \nabla_c Q(\underline{x}|\underline{c}) \right\} = 0 \tag{2.8}$$

We can apply the algorithms (2.3) and (2.6) to (2.8) and functional (2.7) only when the a priori distribution $P(\underline{x})$ is known and, consequently, the mathematical expectation (2.7) can be determined beforehand. Frequently, however, the probability density function $p(\underline{x})$ is unknown. Nonetheless, the optimal vector $\underline{c} = \underline{c}^*$ can still be determined by applying the gradient method using $\nabla_c Q(\underline{x}|\underline{c})$ instead of $E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\}$. This is the advantage of using the method of stochastic approximation. With this method the algorithms for determining $\underline{c} = \underline{c}^*$ can be written in the form

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \nabla_c Q(\underline{x}_j|\underline{c}_j) \tag{2.9}$$

if $Q(\underline{x}|\underline{c})$ is analytic and differentiable, and

$$\underline{c}_{j+1} = \underline{c}_j - \frac{\gamma_j}{2a_j} \left\{ Q_+(\underline{x}_j|\underline{c}_j, a_j) - Q_-(\underline{x}_j|\underline{c}_j, a_j) \right\} \tag{2.10}$$

if $\nabla_c Q(\underline{x}|\underline{c})$ does not exist.

Algorithm (2.9) is a multivariate form of the Robbins-Monro procedure, while algorithm (2.10) is a multivariate form of the Kiefer-Wolfowitz scheme. The analogy between deterministic and stochastic algorithms is apparent. It should be emphasized however, that stochastic algorithms deal with stationary random variables which may contain random noise in addition to the useful signal. The convergence properties of the above algorithms will be considered in the next section.

4. Convergence Properties

In this section the conditions under which the above-mentioned algorithms converge will be described. Since mean square error is used in this report as the only performance criterion, $Q(\underline{x}|\underline{c})$ is analytic and differentiable, and we therefore need to consider only algorithm (2.9).

Let $\underline{c}^*$ satisfy the equation

$$E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\} = 0 \tag{2.11}$$

$E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\}$ is a set of real measurable functions of real variables $\underline{c}$ such that

$$E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\} \begin{cases} > 0 \text{ for } \underline{c} > \underline{c}^* \\ < 0 \text{ for } \underline{c} < \underline{c}^* \\ = 0 \text{ for } \underline{c} = \underline{c}^* \end{cases} \tag{2.12}$$

where $\underline{c} \gtrless \underline{c}^*$ means $c_i \gtrless c_i^*$ for all i.

<u>Theorem</u>: Let $\gamma_1, \gamma_2, \ldots$ be a sequence of positive numbers such that

(A1) $$\lim_{j\to\infty} \gamma_j = 0 \tag{2.13a}$$

(A2) $$\sum_{j=1}^{\infty} \gamma_j = \infty \tag{2.13b}$$

(A3) $$\sum_{j=1}^{\infty} \gamma_j^2 < \infty \tag{2.13c}$$

(A4) $$\gamma_j > 0 \tag{2.13d}$$

Let the following conditions be satisfied

(B) $$\inf_{\varepsilon < \|\underline{c} - \underline{c}^*\| < \frac{1}{\varepsilon}} E\left\{ (\underline{c} - \underline{c}^*)^T \nabla_c Q(\underline{x}|\underline{c})\right\} > 0 \qquad \varepsilon > 0 \tag{2.14}$$

$$\text{(C)} \qquad E\left\{\nabla_c^{\,T} Q(\underline{x}|\underline{c}) \nabla_c Q(\underline{x}|\underline{c})\right\} \leq d(\underline{c}^{*T}\underline{c}^{*} + \underline{c}^{T}\underline{c}) \tag{2.15}$$

for all $\underline{c}$ in a bounded set

and $d > 0$

Then the sequence $\underline{c}_j$ defined by (2.9) converges with probability one to $\underline{c}^*$.[21]

<u>Proof</u>: Subtracting both sides of Eq. (2.9) by $\underline{c}^*$ we have

$$\underline{c}_{j+1} - \underline{c}^* = \underline{c}_j - \underline{c}^* - \gamma_j \underline{\nabla} Q \tag{2.16}$$

where, for simplicity, $Q = Q(\underline{x}|\underline{c})$

Squaring Eq. (2.16)

$$(\underline{c}_{j+1} - \underline{c}^*)^T (\underline{c}_{j+1} - \underline{c}^*) = (\underline{c}_j - \underline{c}^*)^T (\underline{c}_j - \underline{c}^*) - 2\gamma_j (\underline{c}_j - \underline{c}^*)^T \underline{\nabla} Q + \gamma_j^2 \underline{\nabla}^T Q \underline{\nabla} Q$$

and taking the conditional mathematical expectation for given $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_j$, we obtain

$$\begin{aligned} &E\left\{ \|\underline{c}_{j+1} - \underline{c}^*\|^2 \,\middle|\, \underline{c}_1, \underline{c}_2, \ldots, \underline{c}_n \right\} \\ &= \|\underline{c}_j - c^*\|^2 - 2\gamma_j E\left\{ (\underline{c}_j - \underline{c}^*)^T (\nabla Q) \right\} \\ &+ \gamma_j^2 E\left\{ \underline{\nabla}^T Q \underline{\nabla} Q \right\} \end{aligned} \tag{2.17}$$

From condition (C), (2.17) becomes

$$\begin{aligned} E\left\{ \|\underline{c}_{j+1} - \underline{c}^*\|^2 \,\middle|\, \underline{c}_1, \underline{c}_j \right\} &\leq \|\underline{c}_j - \underline{c}^*\|^2 - 2\gamma_j E\left\{ (\underline{c}_j - \underline{c}^*)^T \nabla Q \right\} \\ &+ \gamma_j^2 d(\underline{c}^{*T} \underline{c}^* + \underline{c}_j^T \underline{c}_j) \end{aligned} \tag{2.18}$$

Using condition (B), (2.18) is reduced to

$$E\left\{ \left\| \underline{c}_{j+1} - \underline{c}^* \right\|^2 \Big| \underline{c}_1, \ldots, \underline{c}_j \right\}$$
$$\leq \left\| \underline{c}_j - \underline{c}^* \right\|^2 (1 + \gamma_j^2 d) + 2\gamma_j^2 d\, \underline{c}^T \underline{c}^* \tag{2.18a}$$

Let $$Z_j = \left\| \underline{c}_j - \underline{c}^* \right\|^2 \prod_{k=j}^{\infty} (1+\gamma_k^2 d)$$
$$+ \sum_{k=j}^{\infty} 2d\gamma_k^2 \underline{c}^T \underline{c}^* \sum_{m=k+1}^{\infty} (1 + \gamma_m^2 d) \tag{2.19}$$

Then $$Z_{j+1} = \left\| \underline{c}_{j+1} - \underline{c}^* \right\|^2 \prod_{k=j+1}^{\infty} (1 + \gamma_k^2 d)$$
$$+ \sum_{k=j+1}^{\infty} 2d\gamma_k^2 \underline{c}^T \underline{c}^* \prod_{m=k+1}^{\infty} (1 + \gamma_m^2 d) \tag{2.20}$$

Taking the conditional mathematical expectation for given $\underline{c}_1, \underline{c}_2, \ldots,$ $\underline{c}_j$, we have

$$E\left\{ Z_{j+1} \Big| \underline{c}_1, \ldots, \underline{c}_j \right\} = E\left\{ \left\| \underline{c}_{j+1} - \underline{c}^* \right\|^2 \Big| \underline{c}_1, \ldots \underline{c}_j \right\} \prod_{k=j+1}^{\infty} (1+\gamma_k^2 d)$$
$$+ \sum_{k=j+1}^{\infty} 2d\gamma_k^2 \underline{c}^T \underline{c}^* \prod_{m=k+1}^{\infty} (1+\gamma_m^2 d)$$
$$\leq \left(\left\| \underline{c}_j - \underline{c}^* \right\|^2 (1+d\gamma_j^2) + 2\gamma_j^2 d\underline{c}^T \underline{c}^* \right) \prod_{k=j+1}^{\infty} (1+\gamma_k^2 d)$$
$$+ \sum_{k=j+1}^{\infty} 2d\gamma_k^2 \underline{c}^T \underline{c}^* \prod_{m=k+1}^{\infty} (1+\gamma_m^2 d)$$
$$= Z_j$$

or $$E\left\{ Z_{j+1} \Big| \underline{c}_1, \ldots, \underline{c}_j \right\} \leq Z_j \tag{2.21}$$

Next taking the conditional mathematical expectation for given $\underline{Z}_1, \ldots,$ $\underline{Z}_j$ on both sides of (2.21), we have

$$E\left\{\underline{Z}_{j+1} \middle| \underline{Z}_1, \ldots, \underline{Z}_j\right\} \leq \underline{Z}_j \tag{2.22}$$

Since $\underline{z}_j = f(\underline{c}_1, \underline{c}_2, \cdots, \underline{c}_j)$

Inequality (2.22) shows that $\underline{Z}_j$ is a semimartingale, where

$$E\,\underline{Z}_{j+1} \leq E\,\underline{Z}_j \leq \ldots \leq E\,\underline{Z}_1 < \infty \tag{2.23}$$

so that, according to the theory of semimartingales [22] the sequence $\underline{Z}_1$ converges with probability one, and hence by virture of Eq. (2.19) and (2.13c) the sequence $(\underline{c}_j - \underline{c}^*)$ also converges with probability one to some random number ξ. It remains to show that P $(\xi = 0) = 1$. It is seen that from (2.23), (2.19) and (2.13c) the sequence $E(\underline{c}_j - \underline{c}^*)$ is bounded. Now taking the mathematical expectation on both sides of the inequality (2.18),

$$E\left\{\|\underline{c}_{j+1} - \underline{c}^*\|^2\right\} \leq E\left\{\|\underline{c}_j - \underline{c}^*\|^2\right\} - 2\gamma_j\, E\left\{(\underline{c}_j - \underline{c}^*)^T \nabla Q\right\}$$
$$+\gamma_j^2\, d\left(\underline{c}^{*T}\underline{c}^* + E(\underline{c}_j^T\,\underline{c}_j)\right)$$

and adding the first j inequalities together, we have by deduction

$$E\left\{\|\underline{c}_{j+1} - \underline{c}^*\|^2\right\} \leq E\left\{\|c_1 - c^*\|^2\right\} + \sum_{k=1}^{j}\left(d\,\underline{c}^{*T} c^* \gamma_k^2 + d\gamma_k^2\, E\,(\underline{c}_k^T\,\underline{c}_k)\right)$$
$$- \sum_{k=1}^{j} 2\gamma_k\, E\left\{(\underline{c}_j - \underline{c}^*)^T \nabla Q\right\} \tag{2.24}$$

Since $E\left\{\|\underline{c}_j - \underline{c}^*\|^2\right\}$ is bounded and condition (2.13c) is fulfilled, from Eq. (2.24) it follows that

$$\sum_{k=1}^{\infty} \gamma_k \, E\left\{(c_j - c^*)^T \nabla Q\right\} < \infty \tag{2.25}$$

Using condition (2.13b), i.e., $\sum_{j=1}^{\infty} \gamma_j = \infty$ and noting Eq. (2.14)

$$\inf_{\varepsilon < \|\underline{c} - \underline{c}^*\| < \frac{1}{\varepsilon}} E\left\{(\underline{c} - \underline{c}^*)^T \nabla Q\right\} \geq 0$$

We deduce from (2.25) that

$$E\left\{(\underline{c}_N - \underline{c}^*)^T \nabla Q\right\} \to 0 \text{ with probability one for some sequence N.} \tag{2.26}$$

Now taking $E\left\{\|\underline{c}_j - \underline{c}^*\|^2\right\} \to \underline{\xi}$ with probability 1, and comparing (2.26) with (2.14) we obtain

$$\underline{\xi} = 0 \text{ with probability 1} \tag{2.27}$$

Therefore, algorithm (2.9) converges with probability one

$$P\left\{\lim_{j\to\infty} (\underline{c}_j - \underline{c}^*) = 0\right\} = 1 \tag{2.28}$$

as well as in mean square sense, i.e.,

$$\lim_{j\to\infty} E\left\{\|\underline{c}_j - \underline{c}^*\|^2\right\} = 0 \tag{2.29}$$

5. Geometrical Significances of the Conditions for Convergence

In the last section we mentioned several restrictions imposed on the properties of the sequence $\{\gamma_1, \ldots, \gamma_j\}$ as well as on the behavior of the function $\nabla_c Q(\underline{x}|\underline{c})$. These conditions not only guarantee the convergence of the algorithms but also possess certain geometrical meanings[23].

A. $\gamma_j > 0$. This is to assure that the corrections, on the average, are to be made in the right directions.

B. $\gamma_j \to 0$ as $j \to \infty$. This is to assure that $\underline{c}_j$ calculated from algorithm (2.9) will converge on some specific value. Suppose we let the measured error gradient be $\nabla_c Q(\underline{x}|\underline{c})$ and the real gradient be $E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\}$. Normally there is random noise in measurement

$$\nabla_c Q(\underline{x}|\underline{c}) = E\left\{\nabla_c Q(\underline{x}|\underline{c})\right\} + \xi_j$$

$$j = 1, 2, \ldots$$

Thus $\nabla_c Q(\underline{x}|\underline{c}_j) \nrightarrow 0$ even if $\underline{c}_j = \underline{c}^*$. For $\underline{c}_j$ to converge on any value at all, the condition $\gamma_j \to 0$ as $j \to \infty$ must be satisfied.

It is seen that the method of stochastic approximation is extremely noise resistant. Random independent additive noise ξ_j is eliminated and does not affect the final results.

C. $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$ or $\sum_{j=J}^{j=\infty} \gamma_j^2 \to 0$ as $J \to \infty$. This condition is to account for the accumulative effect of ξ_j. One application of this condition has been seen in the last section. When random noise ξ_j is added at each iteration step, algorithm (2.9) becomes

$$\underline{c}_{j+1} - \underline{c}_j = -\gamma_j \nabla_c Q(\underline{x}|\underline{c}) + \gamma_j \xi_j \tag{2.30}$$

Summing the above equation from $j = J$ upward gives

$$\underline{c}_\infty - \underline{c}_j = -\sum_{j=J}^{\infty} \gamma_j \nabla Q(\underline{x}|\underline{c}) + \sum_{j=J}^{\infty} \gamma_j \xi_j \tag{2.31}$$

(2.31) expresses the total variation in $\underline{c}$ from the Jth step onward.

Since

$$\overline{\left(\sum_{j=J}^{\infty} \gamma_j \xi_j\right)^2} = \overline{\xi^2} \sum_{j=J}^{\infty} \gamma_j^2$$

$\sum_{j=J}^{\infty} \gamma_j^2 \to 0$ assures that the total random variation $\overline{\left(\sum_{j=J}^{\infty} \gamma_j \xi_j\right)^2}$ approaches zero as J becomes very large.

D. $\sum_{j=1}^{\infty} \gamma_j \to \infty$. The above conditions assure that $\underline{c}_j$ converges on some value $\underline{c}_\infty$. $\sum_{j=1}^{\infty} \gamma_j \to \infty$ assures that $\underline{c}_\infty = \underline{c}^*$. Since this condition also implies $\sum_{j=J}^{\infty} \gamma_j \to \infty$, if $\underline{c}_j$ approaches any value other than $\underline{c}^*$, the total correction effect $\sum_{j=J}^{\infty} \gamma_j \nabla_c Q(\underline{x}|\underline{c})$ is infinite. On the other hand, we have no fear of overshoot because each step is very small as $\gamma_j \to 0$ when $j \to \infty$. Conditions A-D state that the rate with which γ_j decreases must be such that, on the one hand, the variance of performance index vanishes, and on the other hand, the variation in γ_j over the variation period is large enough for the law of large numbers to hold.

E. $$\inf_{\varepsilon < ||\underline{c} - \underline{c}^*|| < \frac{1}{\varepsilon}} E\left\{ (\underline{c} - \underline{c}^*)^T \nabla_c Q(\underline{x}|\underline{c}) \right\} \geq 0 \text{ for } \varepsilon > 0.$$

This condition determines the behavior of the surface $E_x\left\{\nabla_c Q(\underline{x}|\underline{c})\right\} = 0$ close to the root and, consequently, the sign of the increments of $\underline{c}_j$. Actually, if the error criterion does have a unique minimum, the above condition is generally satisfied.

F. $$E\left\{\nabla_c^T Q(\underline{x}|\underline{c}) \nabla_c Q(\underline{x}|\underline{c})\right\} \leq d(\underline{c}^{*T}\underline{c}^* + \underline{c}^T\underline{c}) \text{ for } d > 0.$$

This condition requires that the mathematical expectation of the quadratic forms

$$E_x\left\{\nabla_c^T Q(\underline{x}|\underline{c}) \nabla_c Q(\underline{x}|\underline{c})\right\}$$

increase, as $\underline{c}$ increases, no faster than a quadratic paraboloid.

III. TAPPED DELAY LINE FILTERS

1. Optimum Tapped Delay Line Filters

Linear filters can be continuous or discrete. The optimum linear filter developed by Wiener has the form of Eq. (1.5)

$$H_o(\omega) = \frac{\phi_{xd}(\omega)}{\phi_{xx}(\omega)} \tag{3.1}$$

$H_o(\omega)$ may be either physical realizable or not. If it is physically realizable, then standard techniques in network theories can be applied to obtain $H_o(\omega)$ consisting of RLC elements with or without transformers.[24] Another method of synthesizing a continuous linear filter of arbitrary transferfunction (and, hence, impulse response) is to represent it as an infinite linear combination of filters

$$H(\omega) = \sum_{i=1}^{\infty} C_i F_i(\omega) \tag{3.2}$$

where the functions $F_i(\omega)$ are independent and together form a complete set. While an infinite sum is necessary to reproduce exactly the optimum filter response $h(t)$, in practice it might be more useful to find the best filter which can be constructed from a finite number N of such independent components.

One particular type of (3.2) but discrete in nature is the tapped delay line filter. This filter consists of a tapped delay line, or equivalent, with adjustable weights at each tap. In this case

$$H(\omega) = \sum_{k=0}^{N} C_k e^{-j\omega T_k} \tag{3.3}$$

and the impulse response is

$$h(t) = \sum_{k=0}^{N} C_k \delta(t-T_k) \tag{3.4}$$

where $T_k = kT$, T is the delay increment between delay line taps, C_k is

the weight at the kth tap on the filter, and δ is the Dirac delta function. The configuration of such a filter is shown in Fig. 2. D_i denotes a delay of T_i in time.

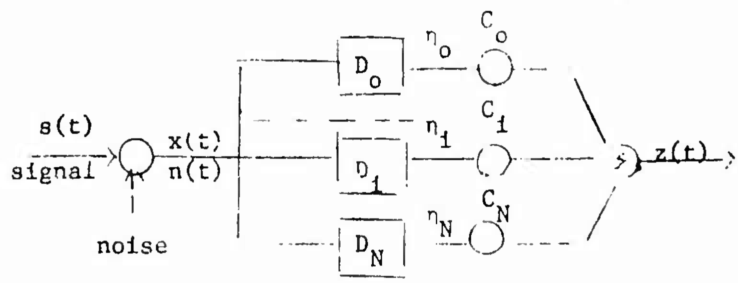


Fig. 2 Tapped Delay Line Filter

The signal η_i obtained at a point after the delay element D_i is of course

$$\eta_i(t) = \int_0^t x(\tau)\ h_i(t-\tau)d\tau$$

$$= \int_0^t x(\tau)\ \delta\ (t- -T_i)d\tau = x(t-T_i). \tag{3.5}$$

The weights $C_i^*(i = 1,2,\ldots,N)$ which optimize any performance criterion may be found by using standard techniques such as calculus of variation or by setting the partial derivatives of the performance criterion with respect to the adjustable gain to zero. Mean squared error criterion $E\{[d(t) - z(t)]^2\}$ is used here because it is simple to use as any and most configurations are not very sensitive to error criterion[25].

A. Frequency domain optimization using calculus of variations.

We are interested in determining $H(\omega)$ minimizing

$$F = E\ \{[d(t) - z(t)]^2\} = \overline{d^2(t)} + \overline{z^2(t)} - 2R_{dz}(o) \tag{3.6}$$

Each term of (3.6) can be related to $H(\omega)$ by means of frequency integral.

Since the filter output is

$$z(t) = \sum_{k=0}^{N} C_k \eta_k(t) \tag{3.7}$$

and its spectral density function is

$$\phi_z(\omega) = \sum_{\ell=0}^{N} \sum_{k=0}^{N} \phi_{z_\ell z_k}(\omega) = \sum_{\ell=0}^{N} \sum_{k=0}^{N} H_\ell(\omega) \; H_k{}^*(\omega) \; \phi_x(\omega)$$

$$= \phi_x(\omega) \sum_{\ell=0}^{N} \sum_{k=0}^{N} C_\ell \; C_k \; e^{-j\omega(\ell-k)T} \tag{3.8}$$

the second term in (3.6) $\overline{z^2(t)}$, is given by

$$\overline{z^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_z(\omega)d\omega = \sum_{\ell=0}^{N} \sum_{k=0}^{N} C_\ell C_k \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_x(\omega) \; e^{-j\omega(\ell-k)T} dw \tag{3.9}$$

The third term is

$$R_{dz}(o) = E \quad d(t) \; Z(t) \quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{dz}(w)dw$$

$$= \sum_{k=0}^{N} \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{d\eta_k}(\omega)d\omega \tag{3.10}$$

and finally

$$\overline{d^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_d(\omega)d\omega \tag{3.11}$$

Combining (3.9), (3.10) and (3.11), we obtain

$$F = \overline{e^2(t)} = \sum_{\ell=0}^{N} \sum_{k=0}^{N} C_k C_\ell \; [\frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_x(\omega) e^{j\omega(\ell-k)T} dw]$$

$$-2 \sum_{\ell=0}^{N} C_k [\frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_d(\omega) e^{j\omega\ell T} d\omega] + \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_d(\omega) d\omega \tag{3.12}$$

The frequency integrals are simply correlation functions; that is,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_x(w)\, e^{j\omega(\ell-k)T} d\omega = R_x\,(\ell T-kT) \tag{3.13}$$

and

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_d(\omega)\, e^{j\omega\ell T} d\omega = R_{dx}(\ell T) \tag{3.14}$$

Now let $c_k = c_k^o + \varepsilon\delta_k,\ k = 0, 1, \ldots, N$

where c_k^o is the optimum and δ_k is an arbitrary constant. For c_k^o optimum, $F(\varepsilon)$ must now have a minimum at $\varepsilon = 0$. Or

$$\left.\frac{dF}{d\varepsilon}\right|_{\varepsilon=0} = \sum_{k=0}^{N}\sum_{\ell=0}^{N} R_x(\ell-k)\Big(\delta_k(c_\ell + \varepsilon\delta_\ell) + \delta_\ell\,(c_k+\varepsilon\delta_k)\Big)$$

$$- 2\sum_{\ell=0}^{N} R_{dx}\,(\ell T)\delta_\ell$$

$$= 2\sum_{\ell=0}^{N}\Big(\sum_{k=0}^{N} c_\ell\, R_x\,(\ell-k) - R_{dx}\,(\ell T)\Big)\delta_\ell = 0$$

for any and all δ_ℓ. Therefore

$$\sum_{k=0}^{N} c_\ell\, R_x(\ell-k) = R_{dx}\,(\ell T) \quad \text{for } \ell = 0, 1, \ldots, N. \tag{3.15}$$

In matrix form

$$\underline{R}_\eta^{\,T}\ \underline{c} = (\underline{R}_{d\eta})$$

or

$$\underline{c}^* = (R_\eta^{\,T})^{-1}\ \underline{R}_{d\eta} \tag{3.16}$$

where

$$\underline{R}_\eta = E\left\{\begin{pmatrix} \eta_0\eta_0 & \cdots & \eta_0\eta_N \\ \cdots & & \cdots \\ \eta_N\eta_0 & \cdots & \eta_N\eta_N \end{pmatrix}\right\} \tag{3.17}$$

and

$$\underline{R}_{d\eta} = E\left\{\left[\begin{pmatrix} d\eta_0 \\ \vdots \\ d\eta_N \end{pmatrix}\right]\right\} \tag{3.18}$$

B. Direct Differentiation

Since $$e(t) = d(t) - \sum_{i=0}^{N} c_i \eta_i(t) = d(t) - \eta^T \underline{c},$$

$$e^2(t) = d^2(t) - 2d\,\eta^T \underline{c} - c^T \eta\, \eta^T c$$

Taking the mathematical expectation, we have

$$\overline{e^2(t)} = \overline{d^2(t)} - 2\overline{d\,\eta}^T \underline{c} + c^T \underline{R}_\eta\, c \tag{3.19}$$

Let $$\nabla_c Q = \begin{pmatrix} \frac{\partial}{\partial c_0} \\ \vdots \\ \frac{\partial}{\partial c_N} \end{pmatrix} Q$$

Since $$\nabla_c \left\{ c^T \underline{R}_\eta\, \underline{c} \right\} = \underline{R}_\eta\, \underline{c} + \underline{R}_\eta^T c$$

together with $\underline{R}_\eta = \underline{R}_\eta^T$

The gradient of $\overline{e^2}$ is

$$\nabla_c \overline{e^2} = -2\,\overline{d\,\eta} + 2\,\underline{R}_\eta^T\, \underline{c} \tag{3.20}$$

For $\overline{e^2}$ to be minimum, we set $\nabla_c \overline{e^2} = 0$.

Then $\underline{c}^* = (\underline{R}_\eta^T)^{-1} \underline{R}_{d\eta}$ same as obtained above.

C. Derivation from Wiener Filter.

Another derivation of $\underline{c}^*$ can be obtained directly from the Wiener filter. From (1.5), the optimum linear filter for additive noise is

$$H_o(\omega) = \frac{\Phi_{dx}(\omega)}{\Phi_{xx}(\omega)} \tag{3.21}$$

Setting $c = c^*$ in Eq. (3.3) and combining with Eq. (3.2) give

$$H_o(\omega) = \sum_{k=0}^{N} c_k^o\, e^{-j\omega kt} = \begin{bmatrix} e^{-j\omega o} \\ e^{-j\omega t} \\ e^{-j\omega nt} \end{bmatrix}^T \left\{ c^* \right\} = \frac{\Phi_{dx}(\omega)}{\Phi_{xx}(\omega)}$$

or

$$\Phi_{xx}(\omega) \begin{bmatrix} e^{-j o t} \\ e^{-j\omega t} \\ \vdots \\ e^{-j\omega nt} \end{bmatrix}^T \left\{ c^* \right\} = \Phi_{dx}(\omega) \tag{3.22}$$

Multiplying both sides of (3.22) by $\left\{ e^{j\omega \ell t} \right\} = \left\{ e^{jo}, \ldots, e^{j\omega nt} \right\}^T$ and integrating from $-\infty$ to $+\infty$, we obtain

$$\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) \left\{ e^{-j\omega kt} \right\}^T \left\{ e^{j\omega \ell t} \right\} d\omega \right) \left\{ c^* \right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{dx}(\omega) \left\{ e^{j\omega \ell t} \right\} d\omega \tag{3.23}$$

Comparing (3.23) with (3.15), we see that they are identical.

2. Minimum Mean Square Error and Effect of Non-optimum Settings.

A. Expressions of mean square error.

The minimum mean square error of the tapped delay line filter is

obtained by substituting the expression of the optimum filter into $\overline{e^2}$.

Using

$$\underline{c}^* = \begin{pmatrix} c_o^* \\ \vdots \\ c_N^* \end{pmatrix} = \underline{R}_\eta^{T\,-1}\, \underline{R}_{d\eta} \tag{3.16}$$

we have

$$\overline{e^2_{min}} = \overline{d^2(t)} - 2\,\overline{d\,\underline{\eta}}^T\, \underline{c}^* + \underline{c}^{*T}\, \underline{R}_\eta\, \underline{c}^*$$

$$= \overline{d^2(t)} - 2\,\overline{d\,\underline{\eta}}^T\, \underline{R}_\eta^{T\,-1}\, \overline{d\,\underline{\eta}}$$

$$+ \overline{d\,\underline{\eta}}^T\, \underline{R}_\eta^{-1}\, \underline{R}_\eta\, \underline{R}_\eta^{T\,-1}\, \overline{d\,\underline{\eta}}$$

$$= \overline{d^2(t)} - \underline{R}_{d\eta}^T\, \underline{R}_\eta^{T\,-1}\, \underline{R}_{d\eta} \tag{3.24a}$$

or

$$= \overline{d^2(t)} - \underline{R}_{d\eta}^T\, \underline{c}^* . \tag{3.24b}$$

or

$$= \overline{d^2(t)} - \overline{z_o^2(t)} \tag{3.24c}$$

where $z_o(t)$ is the output of the optimum filter.

In terms of $\overline{e^2_{min}}$, $\overline{e^2(t)}$ can be expressed as follows:

From $$\overline{e^2(t)} = \overline{d^2(t)} - 2\overline{d\,\underline{\eta}}^T\, \underline{c} + \underline{c}^T\, \underline{R}_\eta\, \underline{c} \tag{3.19}$$

Using Eq. (3.16) and (3.24b), the mean square error is then expressed as

$$\overline{e^2(t)} = \overline{d^2(t)} - 2\overline{d\,\underline{\eta}}^T\, \underline{c} + \underline{c}^T\, \underline{R}_\eta\, \underline{c}$$

$$= \overline{e^2_{min}} + \overline{d\,\underline{\eta}}^T\, \underline{c}^* - 2\overline{d\,\underline{\eta}}^T\, \underline{c} + \underline{c}^T\, \underline{R}_\eta\, \underline{c}$$

$$= \overline{e^2_{min}} + \underline{c}^{*T}\, \underline{R}_\eta\, \underline{c}^* - 2\,\underline{c}^{*T}\, \underline{R}_\eta\, \underline{c} + \underline{c}^T\, \underline{R}_\eta\, \underline{c}$$

$$= \overline{e^2_{min}} + (\underline{c}^T - \underline{c}^{*T})\, \underline{R}_\eta\, (\underline{c} - \underline{c}^*) \tag{3.25}$$

B. Effect on Minimum Mean Square Error due to Non-optimum Settings.

From (3.25) the difference in mean square error due to non-optimum values of $\{c\}$ is

$$\Delta F = \overline{e^2(t)} - \overline{e^2_{min}} = (\underline{c}^T - \underline{c}^{*T})\, R_\eta\, (\underline{c} - \underline{c}^*) \tag{3.26}$$

Let $c_i = c_i^* + \delta_i$

then

$$\Delta F = \underline{\delta}^T\, \underline{R}_\eta\, \underline{\delta}$$
$$= \sum_i \sum_j \delta_i \delta_j \overline{\eta_i \eta_j}$$
$$\leq (N+1)^2 \max_{\text{all } i} |\delta_i| \max_{i,j} |\overline{\eta_i \eta_j}| \tag{3.27}$$

Thus, the error due to non-optimum settings is bounded if the deviations of the weights and the input correlation functions are bounded. Note that for delay line filters $\max_{i,j} |\overline{\eta_i \eta_j}| = R_x(o) = R_s(o) + R_n(o)$.

C. Relationship between the minimum mean square error and the number of time delay elements used in the filter.

From (3.24a) it is seen that

$$\overline{e^2_{min}} = \overline{d^2(t)} - \underline{R}_{d\eta}^{T}\, R_\eta^{T\,-1}\, \underline{R}_{d\eta}$$

Taking the case $d(t) = s(t)$ and noting that

$$\overline{d(t)\,\eta_i(t)} = \overline{s(t)\,x(t-T_i)}$$
$$= \overline{s(t)\{s(t-T_i) + n(t-T_i)\}}$$
$$= \overline{s(t)\,s(t-T_i)} = R_s(T_i),$$

we have

$$\overline{e^2_{min}} = R_s(o) - \left(R_s(o)\ R_s(T_1) \cdot R_s(T_n) \right) \begin{bmatrix} R_x(o) \ldots\ldots R_x(T_n) \\ R_x(T_1)\ R_x(o)\ R_x(T_{n-1}) \\ R_x(T_n) \ldots\ldots\ R_x(o) \end{bmatrix} \begin{bmatrix} R_s(o) \\ R_s(T_1) \\ R_s(T_n) \end{bmatrix}$$

$$\text{where } R_x(T_i) = R_s(T_i) + R_n(T_i) \qquad (3.24d)$$

The last term in the right hand side of (3.24d) is a functional of N, the number of delay line elements, and the correlation functions. For any given forms or values of $R_x(T_i)$ a plot of $\overline{e^2_{min}}$ versus N can be constructed. It is anticipated that the larger N is the smaller $\overline{e^2_{min}}$ will be.

3. Adaptive Tapped Delay Line Filters

The above discu[illegible]s presented a means of determining the optimum values of the gains [illegible]vided that the statistical properties of both the desired signal and the noise are known. Unfortunately, in practice, it is not always possible to know all this information very accurately. If only the filter input and output are available and nothing else, no systematic procedures can be found to adjust the gains. However, if we know something about the system, then we can develop some algorithms to make the filter optimum. It will be shown that if a desired signal is available, or correlation functions of the desired signal, or (not and) correlation functions of the noise can be estimated within acceptable accuracy, the methods of stochastic approximation can be employed to make the filter adaptive to changing operating conditions. These changes may be due to variation in the input signal or the internal structure of the filter. Adaptation is accomplished by observation of the reaction of the

filter to an external signal or to an internal variation with subsequent goal-directed variation of the filter parameters so as to minimize some quality criterion.

The quality criterion may be represented in the form of the mathematical expectation of some strictly convex (not necessarily quadratic) function of the deviation of the output variation from the desired function. For simplicity we shall use the mean squared criterion. Thus,

$$I\,(\underline{c}) = E\left\{Q(d(t) - z(t))\right\} \text{ with } Q(e) = e^2 \tag{3.28}$$

For the tapped delay line filter shown schematically in Fig. 2, we know,

$$x(t) = s(t) + n(t) \tag{1.1}$$

It is assumed here that these functions are stationary random processes. The desired function is the function obtained by applying an arbitrary operation on $s(t)$. This operator may be a differential operator, integral operator, predictor, etc. It can even be a unity operator such that $d(t) = s(t)$. We shall first of all consider the case where $d(t)$ is available. Those cases for which signal or noise correlation functions are known will be treated in a later section. They will turn out to be slight modification of the first case. Nonstationary or time varying systems will be considered subsequently.

For the first case

$$I(\underline{c}) = E\left\{Q(d(t) - z(t))\right\} \tag{3.29}$$

since $$z(t) = \sum_{k=0}^{N} c_k \eta_k(t) = \sum_{k=0}^{N} c_k\, x(t-kT)$$

we have

$$I(\underline{c}) = E\left\{Q(d(t) - \sum_{k=0}^{N} c_k \eta_k(t))\right\}$$

$$= \int_x Q\left(d(t) - \sum_{k=0}^{N} c_k \eta_k(x)\right) P(x)\, dx \tag{3.30}$$

Since $P(x)$ is generally unknown, algorithm (2.9) will be used. For $Q(e) = e^2(t)$, we see that

$$\nabla_c Q(\underline{x}|\underline{c}) = 2e \nabla_c e$$

But

$$\nabla_c e = \nabla_c \left(d(t) - \sum_{k=0}^{N} c_k \eta_k(t)\right)$$

therefore

$$\begin{pmatrix} \frac{\partial e^2}{\partial c_o} \\ \\ \frac{\partial e^2}{\partial c_k} \end{pmatrix} = 2(d(t) - \sum_{k=0}^{N} c_k \eta_k(t)) \begin{pmatrix} -\eta_o(t) \\ \\ -\eta_n(t) \end{pmatrix}$$

and the desired algorithm is

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j e_j \eta_j \tag{3.31}$$

This is precisely the LMS algorithm with constant γ used by Widrow[13] derived from intuitive reasoning rather than from rigorous mathematical proofs.

It would be desirable and instructive to give some physical interpretations of the conditions under which algorithm (3.31) converges.

Algorithm (2.9)

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \nabla_c \, Q(\underline{x}_j | \underline{c}_j) \tag{2.9}$$

converges if the following conditions are satisfied:

$$\text{(a)} \quad \lim_{j \to \infty} \gamma_j = 0, \quad \sum_{j=1}^{\infty} \gamma_j = \infty, \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty \tag{2.13}$$

$$\text{(b)} \quad \inf_{\varepsilon < \|\underline{c} - \underline{c}^*\| \frac{1}{\varepsilon},\ \varepsilon > 0} E\left\{ (\underline{c} - \underline{c}^*)^T \nabla_c \, Q(\underline{x}|\underline{c}) \right\} > 0 \tag{2.14}$$

in the neighborhood of $\underline{c}^*$. $\varepsilon > 0$

$$\text{(c)} \quad E\left\{ \nabla_c^T \, Q(\underline{x}|\underline{c}) \, \nabla_c \, Q(\underline{x}|\underline{c}) \right\} \le d(\underline{c}^{*T}\underline{c}^* + \underline{c}^T\underline{c}). \tag{2.15}$$

The choice of γ_j which satisfies (a) is rather at our own disposal. For example, $\gamma_j = \frac{a}{j+b}$ with a, b > 0 will definitely fulfill the requirement of (a). The remaining conditions depend on the surface of the error gradient, which in turn depends on the choice of error criterion and the physical system under consideration.

Condition (b) is satisfied as long as the function $Q(e)$ is strictly convex. Since $Q(e)$ has a minimum at $\underline{c} = \underline{c}^*$, it is evident that

$$\begin{aligned} \frac{\partial Q}{\partial c_i} &> 0 \text{ for } c_i > c_i^* \\ &= 0 \text{ for } c_i = c_i^* \\ &< 0 \text{ for } c_i < c_i^* \end{aligned} \tag{3.32}$$

$$i = 0, 1, 2, \ldots, N$$

Consequently $\quad (c_i - c_i^*) \dfrac{\partial Q}{\partial c_i} \ge 0$ for all i

and
$$\inf_{\varepsilon < \|\underline{c} - \underline{c}^*\| < \frac{1}{\varepsilon}} E\left\{ (\underline{c} - c^*)^T \nabla_c Q(\underline{x}|\underline{c}) \right\} > 0$$
$$\varepsilon > 0$$

Condition (c) is satisfied if

(a) $\frac{\partial^2 Q}{\partial e^2}$ exists and is uniformly bounded.

(b) $s(t)$ and $n(t)$ are uniformly bounded.

Using a Taylor expansion about $\underline{c} = \underline{c}^*$, we have

$$\frac{\partial Q(\underline{c})}{\partial c_j} = 0 + \sum_{i=0}^{N} (c_i - c_i^*) \left. \frac{\partial^2 Q(\underline{c})}{\partial c_i \partial c_j} \right|_{\underline{c}=\underline{c}^*} \tag{3.33}$$

for arbitrary j, with $j = 0, 1, 2, \ldots, N$.

For the tapped delay line filter with mean square error criterion

$$Q(e) = Q\left(d(t) - \sum_{i=0}^{N} c_i \eta_i(t)\right) \tag{3.34}$$

Therefore,

$$\frac{\partial Q(e)}{\partial c_i} = \frac{\partial Q}{\partial e} (-\eta_i(t))$$
$$\frac{\partial^2 Q(e)}{\partial c_i \partial c_j} = \frac{\partial^2 Q}{\partial e^2} \eta_i(t) \eta_j(t) \tag{3.35}$$

By definition

$$\eta_i(t) = x(t-T_i) = s(t-T_i) + n(t-T_i) \tag{3.36}$$

It is evident that $\frac{\partial^2 Q}{\partial c_i \partial c_j}$ is bounded if conditions (a) and (b) are satisfied.

Therefore,

$$\nabla_c Q(\underline{x}|\underline{c}) \leq k_1 \sum_{i=0}^{N} (c_i - c_i^*) \qquad (3.37)$$

where $k_1 = k \sup_{\text{all } i} \left| \frac{\partial^2 Q(c)}{\partial c_i \partial c_j} \right|$

Taking the inner product and mathematical expectation on each side of (3.37) gives

$$E\left\{\nabla_c^T Q(\underline{x}|\underline{c})\ \nabla_c Q(\underline{x}|\underline{c})\right\}$$

$$\leq k_1^2 \sum_{i=0}^{N} (c_i - c_i^*)^2 \leq k_1^2 \sum_{i=0}^{N} (c_i^2 + c_i^{*2}) \qquad (3.38)$$

$$= d(\underline{c}^{*T}\underline{c}^* + \underline{c}^T\underline{c})$$

In practice conditions (a) and (b) are easily satisfied. Thus the methods of stochastic approximation can be employed in a variety of adaptive processes.

IV. Adaptive Schemes and Rate of Convergence

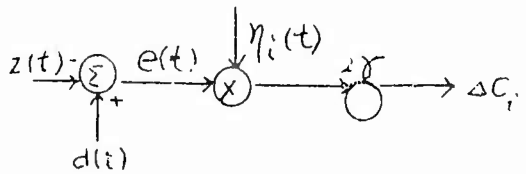
1. Adaptive Schemes

A. An algorithm has been presented to adjust the gains in the tapped delay line filter. If the desired signal is available to generate the error gradient, the adaptive scheme is given by

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \, e_j \, \underline{\eta}_j \tag{4.1}$$

with
$$e(t) = d(t) - \sum_{k=0}^{N} c_k \, \eta_k(t) \tag{4.2}$$

The scheme is shown below



The complete adaptive system is shown in Fig. 3

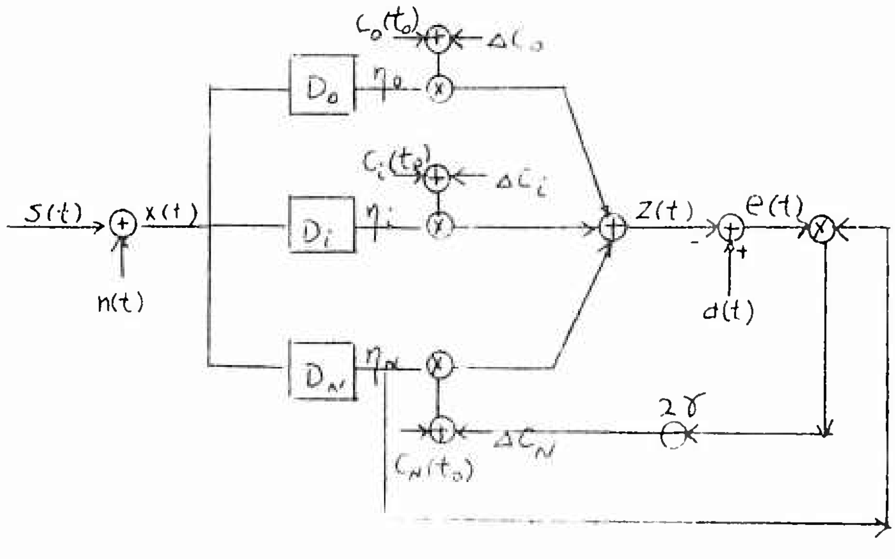


Fig. 3. Adaptive system with d(t) available

B. When $d(t)$ is not available but the statistical properties of the noise are known, the algorithm is modified as follows.

Using $s(t) = x(t) - n(t)$ in the expression of error criterion we have

$$
\begin{aligned}
I(\underline{c}) &= E\left\{ Q(e) \right\} = E\left\{ \left(s(t) - z(t) \right)^2 \right\} \\
&= E\left\{ \left(x(t) - n(t) - z(t) \right)^2 \right\} \\
&= E\left\{ \left(x(t) - z(t) \right)^2 \right\} + E\left\{ n^2(t) \right\} \\
&\quad - 2E\left\{ n(t) \left(x(t) - z(t) \right) \right\}
\end{aligned}
\tag{4.3}
$$

Since

$$
\begin{aligned}
x(t) &= s(t) + n(t) \\
z(t) &= \sum_{k=0}^{N} c_k \eta_k(t) = \sum_{k=0}^{N} c_k x(t - T_k) \\
&= \sum_{k=0}^{N} c_k \left(s(t - T_k) + n(t - T_k) \right)
\end{aligned}
$$

and $\quad E\left\{ s(t)\, n(t) \right\} = 0$

Eq. (4.3) becomes

$$
\begin{aligned}
I(\underline{c}) &= E\left\{ \left(x(t) - z(t) \right)^2 \right\} + E\left\{ n^2(t) \right\} \\
&\quad - 2E\left\{ n^2(t) \right\} + 2E\left\{ n(t) \sum_{k=0}^{N} c_k n(t - T_k) \right\} \\
&= E\left\{ \left(x(t) - z(t) \right)^2 \right\} - E\left\{ n^2(t) \right\} \\
&\quad + 2E\left\{ \sum_{k=0}^{N} c_k n(t) n(t - T_k) \right\} \\
&= E\left\{ \left(x(t) - z(t) \right)^2 \right\} - R_n(o) + 2 \sum_{k=0}^{N} c_k R_n(T_k)
\end{aligned}
\tag{4.4}
$$

where $R_n(T_k)$ is the noise correlation function.

In comparing the algorithm used for the case when $d(t)$ is available, and equal to $s(t)$ if $s(t)$ is replaced by $x(t)$, we would adjust the gains $\underline{c}$ to minimize the

first term on the right hand side of (4.4), i.e., to solve the equation

$$\text{grad } E\left\{\left(x(t) - z(t)\right)^2\right\} = 0$$

Now consider the rest of the right-hand side of (4.4). The second term is independent of $\underline{C}$, and the entries $R_n(T_k)$ appearing in the third term are known. We wish to minimize the sum of the three terms, i.e., find the solution

$$\text{grad } E\left\{\left(x(t) - z(t)\right)^2\right\} + \text{grad } 2 \sum_{k=0}^{N} C_k R_n(T_k) = 0$$

At this point we shall use a modified algorithm whose convergence properties and proofs are found at Appendix A. It is shown that if we let $Q = Q_1 + Q_2$, the algorithm

$$\underline{C}_{j+1} = \underline{C}_j - \gamma_j(\nabla_c Q_1 + \overline{\nabla_c Q_2}) \tag{4.5}$$

also converges in the same sense and under the same physical conditions as algorithm (2.9) for the tapped delay line filters.

In (4.3) we can set

$$Q_1 = [\,x(t) - z(t)]^2$$

$$Q_2 = n^2(t) - 2n(t)\,[x(t) - z(t)]$$

But

$$\overline{Q_2} = -R_n(0) + 2\sum_{k=0}^{N} C_k\, R_n(T_k)$$

and

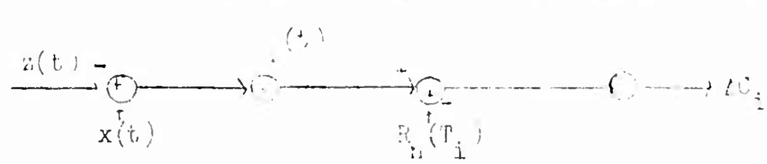
$$\overline{\nabla Q_2} = 2\underline{R}_n = 2[\,R_n(0)\;R_n(T)\;\ldots\;R_n\;(NT)]^T \tag{4.6}$$

Thus, algorithm (4.1) is modified to

$$\underline{C}_{j+1} = \underline{C}_j - \gamma_j\,[\nabla_c\;(x(t) - z(t))^2 + 2\underline{R}_n]^T$$

$$= \underline{C}_j + 2\gamma_j\, n_j(x_j - z_j) - 2\gamma_j \underline{R}_n \tag{4.7}$$

The adaptive scheme is shown below and the whole system is drawn in Fig. 4.



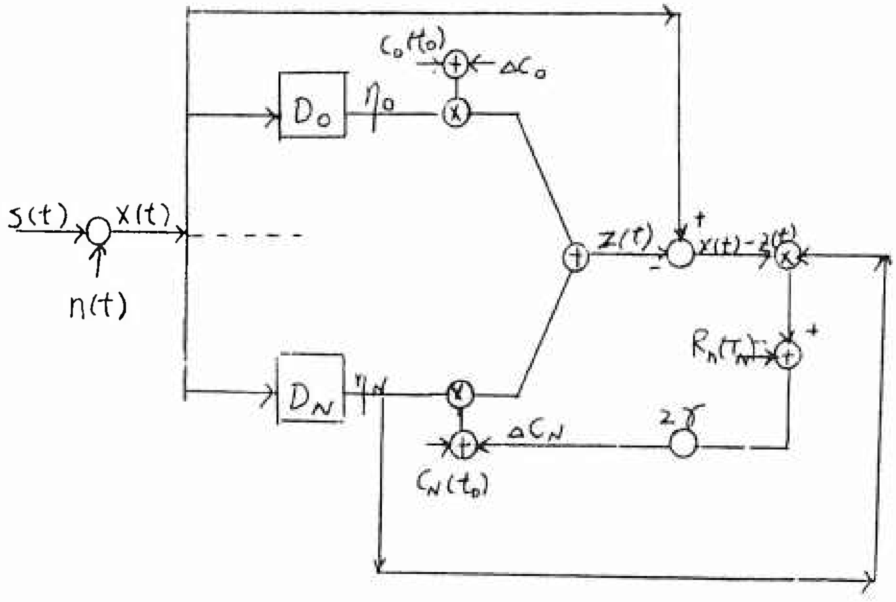


Fig. 4. Adaptive system with known noise statistics

C. Now we shall consider the case when the statistical properties of the signal are known.

Since

$$I(\underline{c}) = E\left\{\left(s(t) - z(t)\right)^2\right\}$$

$$= E\left\{s^2(t)\right\} - 2\,E\left\{s(t)\,z(t)\right\} + E\left\{z^2(t)\right\}$$

$$= E\left\{s^2(t)\right\} - 2\,E\left\{s(t)\sum_{k=0}^{N} c_k\left[s(t-T_k) + n(t-T_k)\right]\right\}$$

$$+ E\left\{\sum_{i=0}^{N}\sum_{j=0}^{N} c_i c_j \eta_i(t)\eta_j(t)\right\}$$

$$= R_s(0) - \sum_{k=0}^{N} 2c_k R_s(T_k) + E\left\{\sum_{i=0}^{N}\sum_{j=0}^{N} c_i c_j \eta_i(t)\eta_j(t)\right\} \qquad (4.8)$$

We can set

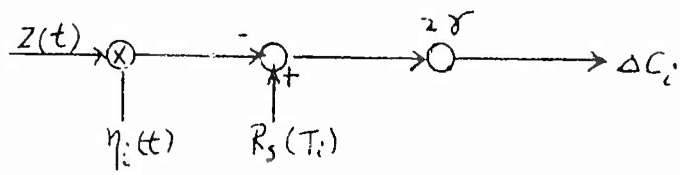
$$Q_1 = + z^2(t) \qquad (4.9)$$

$$Q_2 = s^2(t) - 2s(t)\,z(t) \qquad (4.10)$$

Following the same procedure as before, we have in this case an algorithm

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \underline{R}_S - 2\gamma_j \eta_j z_j \qquad (4.11)$$

the scheme for Eq. (4.11) is then



while the whole system is shown in Fig. 5.

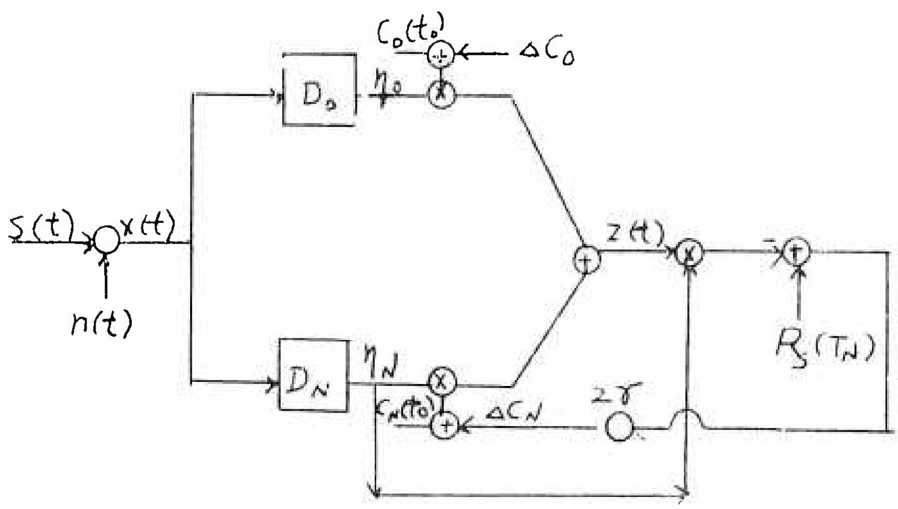


Fig. 5. Adaptive system with known signal statistics.

In the above schemes no distinction between continuous and discrete processes has been made because their connection is obvious. In the discrete case we can set $\gamma_j = \frac{1}{j}$ while its continuous counterpart is $\gamma(t) = \frac{1}{t}$. Theorems concerning the choice of $\gamma(t)$ already exist[26] and are not discussed here.

2. Rate of convergence

Having found an algorithm which converges, we shall investigate how fast it converges. In other words, we would like to know the mean square error at each stage during the adaptation period.

From (4.1)

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \underline{\eta}_j e_j$$

and using the expression

$$e_j = d_j - z_j = d_j - \sum_{i=0}^{N} c_j^i \eta_j^i$$

$$= d_j - \underline{\eta}_j^T \underline{c}_j$$

we get the corresponding matrix form

$$\underline{c}_{j+1} = \underline{c}_j - 2\gamma_j \underline{\eta}_j \underline{\eta}_j^T \underline{c}_j + 2\gamma \underline{d} \underline{\eta}_j$$

$$= (1 - 2\gamma_j \underline{\eta}_j \underline{\eta}_j^T) \underline{c}_j + 2\gamma_j c \underline{\eta}_j \qquad (4.12)$$

Taking the mathematical expectation of Eq. (4.12) and diagonizing the matrix $E\left\{\underline{\eta}_j \underline{\eta}_j^T\right\}$ such that

$$E\left\{\underline{\eta}_j \underline{\eta}_j^T\right\} = \underline{R}_\eta = \underline{P}^{-1} \underline{\Lambda} \underline{P}$$

where $\underline{P}$ is an orthonormal matrix, and $\underline{\Lambda} = \begin{pmatrix} \lambda_0 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix}$ is the eigenvalue matrix, we obtain

$$\overline{\underline{c}}_{j+1} = (1 - 2\gamma_j \underline{R}_\eta) \overline{\underline{c}}_j + 2\gamma_j \overline{d_j \underline{\eta}_j}$$

$$= (1 - 2\gamma_j \underline{P}^{-1} \underline{\Lambda} \underline{P}) \overline{\underline{c}}_j + 2\gamma_j \overline{d_j \underline{\eta}_j} \qquad (4.13)$$

AVERAGING PROCESS

In the above we assumed that $\underline{c}$ is statistically independent of $\underline{\eta}$. Although $\underline{c}$ can not affect $\underline{\eta}$ in any manner, the increment of $\underline{c}$ at each

stage is, however, related to $\underline{n}$ by (4.1). Since the increment is generally very small and the total effect involves addition of a large number of small increments, we can assume $\overline{\underline{c}\,\underline{\eta}} = \overline{\underline{c}}\,\overline{\underline{\eta}}$ in a manner similar to that used in the analysis of phase-locked loops*

Let us define

$$\overline{\underline{W}} = \underline{P}\,\underline{C}, \qquad \underline{n}' = \underline{P}\ \underline{n} \tag{4.14}$$

then (4.13) becomes

$$\overline{\underline{W}}_{j+1} = (1 - 2\gamma_j \Lambda)\,\overline{\underline{W}}_j + 2\gamma_j \ \overline{d\underline{n}'} \tag{4.15}$$

Since $\overline{d\underline{n}'} = \underline{R}_n\,\underline{C}^*$ as seen from (3.16), we have

$$\overline{\underline{W}}_{j+1} - \underline{W}^* = (1 - 2\gamma_j \Lambda)\,(\overline{\underline{W}}_j - \underline{W}^*) \tag{4.16}$$

Now consider any particular component w of $\underline{W}$ and for clarity no subscript or supuscript indicating the component is used. Then

$$\overline{W}_{j+1} - W^* = (1 - 2\gamma_j \lambda)(\overline{W}_j - W^*) \tag{4.17}$$

Using Eq. (4.17) recursively gives

$$\overline{W}_j = (W_1 - W^*) \prod_{k=1}^{j-1} (1 - 2\gamma_j \lambda) + W^* \tag{4.18}$$

* Viterbi, A.J., <u>Principles of Coherent Communication</u>, McGraw Hill Book Co. New York, 1966.

We shall now find $\overline{w_j^2}$.

From (4.1) and taking the product of $\underline{c}$ and $\underline{c}^T$, we obtain

$$\underline{c}_{j+1}\ \underline{c}_{j+1}^T = (\underline{c}_j + 2\gamma_j\ e_j\ \underline{n}_j)\ (\underline{c}_j^T + 2\gamma_j\ e_j\ \underline{n}_j^T)$$

$$= \underline{c}_j\ \underline{c}_j^T + 2\gamma_j\ e_j\ (\underline{c}_j\ \underline{n}_j^T + \underline{n}_j\ \underline{c}_j^T) + 4\gamma_j^2\ e_j^2\ \underline{n}_j\ \underline{n}_j^T \tag{4.19}$$

Since

$$e_j(\underline{c}_j\ \underline{n}_j^T + \underline{n}_j\ \underline{c}_j^T)$$

$$= (d_j - \underline{n}_j^T\ \underline{c}_j)\ (\underline{c}_j\ \underline{n}_j^T + \underline{n}_j\ \underline{c}_j^T)$$

$$= \underline{c}_j\ d_j\underline{n}_j^T + d_j\ \underline{n}_j\ \underline{c}_j^T$$

$$- \underline{c}_j\ \underline{c}_j^T\ n_j\ \underline{n}_j^T - \underline{n}_j\ \underline{n}_j^T\ \underline{c}_j\ \underline{c}_j^T$$

Note $\quad \underline{A}\ \underline{B}^T + \underline{B}\ \underline{A}^T = 2\ \{\underline{A}\ \underline{B}^T\}^s$

where s denotes the symmetrical part of a matrix. For example, if

$$\underline{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then

$$\underline{A}^s = \begin{bmatrix} \frac{1}{2}\ (a_{11} + a_{11}) & \frac{1}{2}\ (a_{12} + a_{21}) \\ \frac{1}{2}\ (a_{21} + a_{12}) & \frac{1}{2}\ (a_{22} + a_{22}) \end{bmatrix}$$

we have

$$\overline{e_j\ (\underline{c}_j\ \underline{n}_j^T + \underline{n}_j\ \underline{c}_j^T)}$$

$$= 2\left\{R_\eta\ \overline{c^*\ c_j^T}\right\}^s - 2\ \left\{R_\eta\ \overline{c_j\ c_j^T}\right\}^s$$

$$= 2\left\{\underline{R}_\eta(\underline{c}^* - \underline{c}_j)\ \underline{c}_j^T\right\}^s \tag{4.20}$$

Taking mathematical expectation on both side of (4.19 and using (4.20) yield

$$\overline{\underline{c}_{j+1}\,\underline{c}_{j+1}^T} = \overline{\underline{c}_j\,\underline{c}_j^T} + 4\gamma_j \left\{ \overline{\underline{R}_\eta\,(\underline{c}^* - \underline{c}_j)\,\underline{c}_j^T} \right\}^s$$

$$+ \; 4\gamma_j^{\,2}\;\overline{e_j^{\,2}\,\underline{\eta}_j\,\underline{\eta}_j^T} \qquad (4.21)$$

For large j , the following approximation can be made

$$\overline{e_j^{\,2}\;\underline{\eta}_j\,\underline{\eta}_j^T} \simeq \overline{e^2_{min}\;\underline{\eta}_j\,\underline{\eta}_j^T} = \overline{e^2_{min}}\;\underline{R}_\eta \qquad (4.22)$$

(4.22) can be viewed as a Taylor series expansion around the optimum point and with higher order terms neglected for large j .

Therefore, (4.21) becomes

$$\overline{\underline{c}_{j+1}\,\underline{c}_{j+1}^T} = \overline{\underline{c}_j\,\underline{c}_j^T} + 4\gamma_j \left\{ \overline{\underline{R}_\eta\,(\underline{c}^* - \underline{c})\,\underline{c}^T} \right\}_j^s$$

$$+ \; 4\gamma_j^{\,2}\;\overline{e^2_{min}}\;\underline{R}_\eta \qquad (4.23)$$

Using the transformation $\underline{c} = \underline{P}^{-1}\,\underline{W}$ as defined in (4.14), we can change (4.23) to the form

$$\overline{\underline{W}_{j+1}\,\underline{W}_{j+1}^T} = \overline{\underline{W}_j\,\underline{W}_j^T} + 4\gamma_j\underline{P} \left\{ \overline{\underline{P}^{-1}\,\underline{\Lambda}\,\underline{P}\,\underline{P}^{-1}\,(\underline{W}^* - \underline{W})\underline{W}^T\underline{P}} \right\}_j^s \underline{P}^{-1}$$

$$+ \; 4\gamma_j^{\,2}\;\underline{P}\;\overline{e^2_{min}}\;\underline{P}^{-1}\,\underline{\Lambda}\,\underline{P}\,\underline{P}^{-1}$$

And

$$\left\{ \overline{\underline{W}_{j+1}\,\underline{W}_{j+1}^T} \right\}^D = \left\{ \overline{\underline{W}_j\,\underline{W}_j^T} \right\}^D + 4\gamma_j \left\{ \overline{\underline{\Lambda}\,(\underline{W}^* - \underline{W})\,\underline{W}^T} \right\}_j^D$$

$$+ \; 4\gamma_j^{\,2}\;\overline{e^2_{min}}\;\underline{\Lambda} \qquad (4.24)$$

In the above D denotes the diagonal elements of a matrix. These elements have

the desired form $\overline{w^2}$, which can be expressed as

$$\overline{w_{j+1}^2} = \overline{w_j^2} + 4\gamma_j \lambda \overline{(w^* - w_j) w_j} + 4\gamma_j^2 \lambda \overline{e_{min}^2} \tag{4.25}$$

$$= (1 - 4\gamma_j\lambda) \overline{w_j^2} + 4\gamma_j\lambda w^* \bar{w}_j + 4\gamma_j^2\lambda \overline{e_{min}^2}$$

From (4.16) we have

$$\overline{W_{j+1} W^T_{j+1}} = (1 - 4\gamma_j \underline{\Lambda}) \left[\bar{W}_j \bar{W}_j^T\right]^s + 4\gamma_j \underline{\bar{W}^*} \underline{\bar{W}}_j^T \tag{4.26}$$

Let $\theta_j = \overline{(W_j - \bar{W}_j)(W_j - \bar{W}_j)^T} = (\overline{W W^T})_j - \bar{W}_j \bar{W}_j^T$

Substrating the diagonal terms of (4.24) from those of (4.26), we obtain

$$\underline{\theta}_{j+1}^D = (1 - 4\gamma_j \quad) \underline{\theta}_j^D + 4\gamma_j^2 \underline{\Lambda} \overline{e_{min}^2} \tag{4.27}$$

Since $\underline{\theta}_j^D$ has the elements $\overline{w_j^2} - \bar{w}_j^2$, we see that for any particular component of θ_j^D,

$$\overline{w_{j+1}^2} - \bar{w}_{j+1}^2 = \theta_{j+1} = (1 - 4\gamma_j\lambda) \theta_j + 4\gamma_j^2\lambda \overline{e_{min}^2} \tag{4.28}$$

Iterating backward,

$$\theta_{j+1} = \theta_1 \prod_{k=1}^{j} (1 - 4\gamma_k\lambda)$$

$$+ 4\lambda\overline{e_{min}^2} \sum_{k=1}^{j} \gamma_k^2 \prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell\lambda)$$

But $\theta_1 = 0$ because $\bar{w}_1 = w_1$,

$$\theta_{j+1} = 4\lambda \overline{e_{min}^2} \sum_{k=1}^{j} \gamma_k^2 \prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell\lambda) \tag{4.29}$$

or

$$\overline{(w_{j+1} - \bar{w}_{j+1})^2} = 4\gamma\overline{e_{min}^2} \sum_{k=1}^{j} \gamma_k^2 \prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell\lambda) \tag{4.30}$$

Several special cases will be considered.

(1) Setting $\gamma_j = \dfrac{1}{2(j+1)\lambda}$ (4.31)

This is a legitimate expression as γ_j defined by (4.31) satisfies all the required conditions for convergence.

Note
$$\prod_{k=1}^{j} \left(1 - \frac{1}{k+1}\right) = \prod_{k=1}^{j} \frac{k}{k+1} = \frac{1}{j+1} \tag{4.32}$$

(4.18) gives us

$$\bar{w}_{j+1} = \frac{1}{j+1} (w_1 - w^*) + w^* \tag{4.33}$$

Note also that [see Eq.(B.8) of Appendix B],

$$\prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell \lambda) = \prod_{\ell=k+1}^{j} \left(1 - \frac{2}{\ell+1}\right) \simeq \frac{(k+1)^2}{(j+1)^2} \tag{4.34}$$

$$\sum_{k=1}^{j} \gamma_k^2 \prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell \lambda) = \sum_{k=1}^{j} \frac{1}{4\lambda^2 (k+1)^2} \frac{(k+1)^2}{(j+1)^2}$$

$$= \frac{1}{4\lambda^2} \sum_{k=1}^{j} \frac{1}{(j+1)^2} = \frac{1}{4\lambda^2} \frac{j}{(j+1)^2} \tag{4.35}$$

thus (4.30) gives us

$$\overline{(w_{j+1} - \bar{w}_{j+1})^2} = \frac{\overline{e^2_{min}}}{\lambda} \frac{j}{(j+1)^2} \tag{4.36}$$

As derived in (3.25), the mean squared error at any time is given by

$$\overline{e_j^2} = \overline{e^2_{min}} + (\underline{c}_j - \underline{c}^*)^T R_n (\underline{c}_j - \underline{c}^*)$$

$$= \overline{e^2_{min}} + (\underline{c}_j - \underline{c}^*)^T \underline{P}^{-1} \underline{\Lambda} \underline{P} (\underline{c}_j - \underline{c}^*)$$

$$= \overline{e^2_{min}} + (\bar{\underline{W}}_j - \bar{\underline{W}}^*)^T \underline{\Lambda} (\bar{\underline{W}}_j - \bar{\underline{W}}^*) \tag{4.37}$$

The expected difference between the mean squared error at each stage during the adaptation period and the minimum mean squared error is then

$$E\left\{\overline{e^2_{j+1}} - \overline{e^2_{min}}\right\} = E\left\{(W_{j+1} - W^*)^T \Lambda (\bar{W}_{j+1} - \underline{W}^*)\right\}$$

$$= E\left\{\sum_{i=0}^{N} \lambda_i (w_{j+1,i} - w^*_i)^2\right\}$$

$$= \sum_{i=0}^{N} \lambda_i \overline{(w_{j+1,i} - w^*_i)^2} \qquad (4.38)$$

But

$$\overline{(w_{j+1} - w^*)^2} = \overline{(w_{j+1} - \bar{w}_{j+1})^2} + \bar{w}_{j+1}^2 - 2 w^* \bar{w}_{j+1} + w^{*2}$$

Using (4.33) and (4.36), we have

$$\overline{(w_{j+1} - w^*)^2} = \frac{\overline{e^2_{min}}}{\lambda} \frac{j}{(j+1)^2} + \frac{1}{(j+1)^2} (w_1 - w^*)^2 \qquad (4.39)$$

(4.38) becomes

$$E\left\{\overline{e^2_{j+1}} - \overline{e^2_{min}}\right\} = \frac{j}{(j+i)^2} \sum_{k=0} \overline{e^2_{min}} + \frac{1}{(j+1)^2} \sum_{k=0}^{N} \lambda_k (w_{1,k} - w^*_k)^2$$

$$= \frac{j}{(j+1)^2} (N + 1) \overline{e^2_{min}} + \frac{1}{(j+1)^2} (\underline{c}_1 - \underline{c}^*)^T \underline{R}_\eta (\underline{c}_1 - \underline{c}^*) \qquad (4.40)$$

The last step is obtained from

$$\sum_{k=0}^{N} \lambda_i w_i^2 = \underline{W}^T \underline{\Lambda} \underline{w} = \underline{c}^T \underline{R}_\eta \underline{c}$$

Thus for large j ,

$$E\left\{\overline{e^2_{j+1}} - \overline{e^2_{min}}\right\} \cong \frac{(N+1) \overline{e^2_{min}}}{j+1} \qquad (4.41)$$

(4.40) is the desired expression for the rate of convergence. For large j , the mean squared error decreases approximately as the first power of time.

(2) Setting $\gamma_j = \dfrac{1}{2(j+1)}$. (4.42)

The choice of γ_j defined by (4.31) requires some a priori knowledge about the signal and noise properties. Otherwise, if the correlation matrix $\underline{R}_n$ is not known, the eigenvalues λ_i cannot be determined. The arbitrary choice of λ_j defined by (4.42) will be studied.

From (4.18) with $\gamma_j = \dfrac{1}{2(j+1)}$ we have

$$w_{j+1} = (w_1 - w^*) \prod_{k=1}^{j} \left(1 - \frac{\lambda}{j+1}\right) + w^* \tag{4.43}$$

But

$$\prod_{k=1}^{j} \left(1 - \frac{\lambda}{j+1}\right) = \frac{\Gamma(j+1-\lambda)}{(j+1)!\,\Gamma(2-\lambda)}$$

$$\simeq \frac{1}{\Gamma(2-\lambda)(j+1)^{\lambda}} \quad \text{for } j >> 1 \quad \text{and} \quad j >> \lambda \tag{4.44*}$$

Thus

$$\bar{w}_{j+1} = \frac{(w_1 - w^*)}{\Gamma(2-\lambda)(j+1)^{\lambda}} + w^* \tag{4.45}$$

Note also that

$$\prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell \lambda) = \prod_{\ell=k+1}^{j} \left(1 - \frac{2\lambda}{\ell+1}\right) \simeq \frac{(k+1)^{2\lambda}}{(j+1)^{2\lambda}} \tag{4.46}$$

Therefore,

$$\sum_{k=1}^{j} \gamma_k^{\,2} \prod_{\ell=k+1}^{j} (1 - 4\gamma_\ell \lambda) = \sum_{k=1}^{j} \frac{1}{4(k+1)^2} \; \frac{(k+1)^{2\lambda}}{(j+1)^{2\lambda}}$$

$$= \frac{1}{4(j+1)^{2\lambda}} \sum_{k=1}^{j} (k+1)^{2\lambda-2}$$

Using the formula (No. 29.9, Tables of Integrals by Dwight)

$$\sum_{u=1}^{n} u^{p} = \frac{n^{p+1}}{p+1} + \frac{n^{p}}{2} + \frac{1}{12}\, p\, n^{p-1} - \frac{1}{30}\,\frac{1}{4!}\, p(p-1)(p-2)n^{p-3} + \cdots$$

---

* Derivation appears in Appendix B.

we can let $k+1 = n$, $p = 2\lambda - 2$, $n = j+1$, and obtain

$$\sum_{k=1}^{j} (k+1)^{2\lambda-2} = \sum_{u=2}^{j+1} u^{p} = \sum_{u=1}^{j+1} u^{p} - 1$$

$$= -1 + \frac{(j+1)^{2\lambda-1}}{2\lambda-2} + \frac{(j+1)^{2\lambda-3}}{2} + \frac{1}{12} (2\lambda-2)\ (j+1)^{2\lambda-3} + --- \qquad (4.48)$$

(4.47) then becomes

$$\sum_{k=1}^{j} \gamma_k^{2} \prod_{\ell=k+1}^{j} (1-4\gamma_\ell\lambda) = \frac{1}{4(j+1)^{2\lambda}} [-1 + \frac{(j+1)^{2\lambda-1}}{2\lambda-2} - \frac{(j+1)^{2\lambda-2}}{2} + ---]$$

$$\approx \frac{1}{4} [\frac{-1}{(j+1)^{2\lambda}} + \frac{1}{(2\lambda-2)} \frac{1}{j+1}] \text{ for large } j . \qquad (4.49)$$

Substituting (4.49) into (4.30) and combining with (4.43) yield

$$\overline{(w_{j+1} - w^*)^2} \approx \frac{\lambda\ \overline{e^2_{min}}}{2\lambda - 2} \frac{1}{j+1} + \frac{1}{(j+1)^{2\lambda}} \left(\frac{(w_1 - w^*)^2}{\Gamma^2(2-\lambda\ )} - \lambda\ \overline{e^2_{min}} \right) \qquad (4.50)$$

and

$$E \left\{ \overline{e^2_{j+1}} - \overline{e^2_{min}} \right\} = \sum_{k=0}^{N} \lambda_k \overline{(w_{j+1,k} - w^*_k)^2}$$

$$\approx \sum_{k=0}^{N} \frac{\lambda_k^2 \overline{e^2_{min}}}{(j+1)\ (2\lambda_k-2)} + \frac{\lambda_k}{(j+1)^{2\lambda_k}} [\frac{(w_{1,k} - w^*_k)^2}{\Gamma^2(2-2\lambda_k)} - \lambda_k \overline{e^2_{min}}] \qquad (4.51)$$

(3) $\gamma_j = \gamma =$ constant

The expressions for γ_j defined by (4.31) and (4.42) satisfy the conditions for convergence as stated in (2.13). In these cases the γ_j and thus the gain increment Δc_j become smaller and smaller as time j proceeds during the adaptation period. It is anticipated that the rate of convergence will be increased if a small constant value is set for γ . As shown by Comer[29] , the algorithm with constant γ_o has comparatively little noise resistance. Furthermore, in the presence of measuring error with variance σ^2 , convergence in the usual sense

does not occur, but

$$\lim_{j\to\infty} E\left\{ \| C_j - C^* \|^2 \right\} < F(\gamma_o , \sigma^2) \tag{4.53}$$

and

$$F(\gamma_o , \sigma^2) \to o \quad \text{as} \quad \gamma_o \to o$$

Now we shall study the rate of convergence when γ is a constant. From Eq.(4.18) we see that with $\gamma_j = \gamma = \text{const}$,

$$\bar{w}_{j+1} = (w_1 - w^*) \prod_{k=1}^{j} (1 - 2\gamma\lambda) + w^*$$

$$= (1 - 2\gamma\lambda)^j (w_1 - w^*) + w^* \tag{4.54}$$

Since

$$a + a\gamma + a\gamma^2 + \dots + a\gamma^{n-1} = \frac{a(1 - \gamma^n)}{1 - \gamma}$$

We can obtain

$$\sum_{k=1}^{j} (1 - 4\gamma\lambda)^{-k} = \frac{1}{4\gamma\lambda} [(1 - 4\gamma\lambda)^{-(j-1)} - 1]$$

Thus

$$\sum_{k=1}^{j} \gamma_k^2 \prod_{\ell=k+1}^{j} (1 - 4\gamma\lambda) = \sum_{k=1}^{j} \gamma^2 (1 - 4\gamma\lambda)^{j-k-1}$$

$$= \gamma^2 (1 - 4\gamma\lambda)^{j-1} \sum_{k=1}^{j} (1 - 4\gamma\lambda)^{-k}$$

$$= \gamma^2 \frac{1}{4\gamma\lambda} [1 - 4\gamma\lambda)^{j-1}]$$

and (4.30) becomes

$$\overline{(w_{j+1} - \bar{w}_{j+1})^2} = \overline{e^2_{min}} \gamma [1-(1-4\gamma\lambda)^{j-1}]$$

The mean squared error is then

$$E\left\{\overline{e^2_{j+1}} - \overline{e^2_{min}}\right\} = \overline{e^2_{min}}\ \gamma \sum_{k=0}^{N} \lambda_k\ [1-(1-4\gamma\lambda)^{j-1}]$$

$$+ \sum_{k=0}^{N} \lambda_k\ (w_{1,k} - w^*_k)^2\ (1 - 2\gamma\lambda_k)^{2j} \tag{4.55}$$

It is seen from (4.35) that if the error is to decrease at all, one basic requirement should be met, i.e.,

$$0 < 1 - 4\gamma\lambda < 1 \quad \text{with} \quad \gamma > 0 \tag{4.56}$$

which implies

$$0 < \gamma < \frac{1}{4\lambda_{max}} \tag{4.57}$$

λ_{max} is the largest eigenvalue of the correlation matrix $\underline{R}_n$. Thus γ = constant cannot be set at will if st ility of the adaptive loop is to be maintained.

The rate of convergence has been obtained so far only for the algorithm with the availability of a desired signal to generate the real time error function $e(t)$. Now we shall compare the algorithm

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j\ e_j\ \underline{n}_j$$

$$= \underline{c}_j + 2\gamma_j\ s_j\ \underline{n}_j - 2\gamma_j\ z_j\ \underline{n}_j \tag{4.58}$$

where s_j replaces d_j for the desired signal with the other two

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j\ \underline{R}_s - 2\gamma_j\ z_j\ \underline{n}_j \tag{4.59}$$

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j\ \underline{n}_j\ (x_j - z_j) - 2\gamma_j\ \underline{R}_n \tag{4.60}$$

when signal or noise correlation functions are used.

Taking mathematical expectation on both sides of (4.58) gives

$$\bar{\underline{c}}_{j+1} = \bar{\underline{c}}_j + 2\gamma_j \overline{s_j \underline{n}_j} - 2\gamma_j \overline{z_j \underline{n}_j} \tag{4.61}$$

But

$$\overline{s\eta} = E\left\{ s(t) \begin{bmatrix} s(t) + n(t) \\ s(t-T) + n(t-T) \\ s(t-NT) + n(t-NT) \end{bmatrix} \right\}$$

$$= \begin{bmatrix} R_s(o) \\ R_s(T) \\ \vdots \\ R_s(NT) \end{bmatrix} \triangleq \underline{R}_s \tag{4.62}$$

and

$$\overline{\underline{z}_j \underline{n}_j} = \overline{\underline{n}_j \underline{n}_j^T \underline{c}_j} = \underline{R}_n \bar{\underline{c}}_j \tag{4.63}$$

we thus have

$$\bar{\underline{c}}_{j+1} = (1 - 2\gamma_j \underline{R}_n) \bar{\underline{c}}_j + 2\gamma_j \underline{R}_s \tag{4.64}$$

Taking the average on both sides of (4.59) gives

$$\bar{\underline{c}}_{j+1} = \bar{\underline{c}}_j + 2\gamma_j \underline{R}_s - 2\gamma_j \overline{z_j \underline{n}_j}$$

which is identical to (4.64) by virtue of (4.62).

Taking the average on both sides of (4.60) gives

$$\bar{\underline{c}}_{j+1} = \bar{\underline{c}}_j + 2\gamma_j \overline{\underline{n}_j (x_j - z_j)} - 2\gamma_j \underline{R}_n \tag{4.65}$$

But

$$\overline{n_j x_j} - R_n = E\left\{ \begin{bmatrix} s(t) + n(t) \\ s(t-T) + n(t-T) \\ s(t-NT) + n(t-NT) \end{bmatrix} \right\} [s(t) + n(t)] \quad - R_n$$

$$= \begin{bmatrix} R_s(o) + R_n(o) \\ R_s(T) + R_n(T) \\ -- \\ R_s(NT) + R_n(NT) \end{bmatrix} - \begin{bmatrix} R_n(o) \\ R_n(T) \\ -- \\ R_n(NT) \end{bmatrix} = \begin{bmatrix} R_s(o) \\ R_s(T) \\ \\ R_s(NT) \end{bmatrix} = \underline{R}_s \tag{4.66}$$

(4.65) can then be reduced to (4.64).

However, (4.64) is just (4.13) if d(t) is replaced by s(t). We therefore can conclude that for filtering problem where $d(t) = s(t)$, the expected values for the gains at any stage are given by the same formula, i.e.,

$$\overline{w}_{j+1} = (w_1 - w^*) \prod_{k=1}^{j} (1 - 2\gamma_j \lambda) + w^* \tag{4.67}$$

which is valid for the transformed gain components.

Let us now consider the variation of $\overline{c_j\, c_j^T}$ for the other two cases. Taking the product of each side with its transpose in (4.59) gives

$$\begin{aligned}
\underline{c}_{j+1}\, \underline{c}_{j+1}^T = {} & \underline{c}_j\, \underline{c}_j^T + 2\gamma_j\, \underline{c}_j\, \underline{R}_s^T - 2\gamma_j\, z_j\, \underline{c}_j\, \underline{n}_j^T \\
& + 2\gamma_j\, \underline{R}_s\, \underline{c}_j^T + 4\gamma_j^2\, \underline{R}_s\, \underline{R}_s^T - 4\gamma_j^2\, z_j\, [R_s]\, [n]_j^T \\
& - 2\gamma_j\, z_j\, \underline{n}_j\, \underline{c}_j^T - 4\gamma_j^2\, z_j\, \underline{n}_j\, \underline{R}_s^T + 4\gamma_j^2\, z_j^2\, \underline{n}_j\, \underline{n}_j^T
\end{aligned}$$

$$\begin{aligned}
= {} & \underline{c}_j\, \underline{c}_j + 2\gamma_j\, (\underline{R}_s\, \underline{c}_j^T + \underline{c}_j\, \underline{R}_s^T) \\
& - 2\gamma_j\, z_j\, (\underline{c}_j\, \underline{n}_j^T + \underline{n}_j\, \underline{c}_j^T) \\
& - 4\gamma_j^2\, z_j\, (\underline{R}_s\, \underline{n}_j^T + \underline{n}_j\, \underline{R}_s^T) \\
& + 4\gamma_j^2\, \underline{R}_s\, \underline{R}_s^T + 4\gamma_j^2\, z_j^2\, \underline{n}_j\, \underline{n}_j^T
\end{aligned}$$

$$\begin{aligned}
= {} & \underline{c}_j\, \underline{c}_j^T + 4\gamma_j\, (\underline{R}_s\, \underline{c}_j^T)^s - 4\gamma_j\, (\underline{n}_j\, \underline{n}_j^T\, \underline{c}_j\, \underline{c}_j^T)^s \\
& - 8\gamma_j^2\, (\underline{R}_s\, \underline{c}_j^T\, \underline{n}_j\, \underline{n}_j^T)^s + 4\gamma_j^2\, \underline{R}_s\, \underline{R}_s^T + 4\gamma_j^2\, z_j^2\, \underline{n}_j\, \underline{n}_j^T
\end{aligned}$$

When average is taken on both sides, we have

$$\begin{aligned}
\overline{\underline{c}_{j+1}\, \underline{c}_{j+1}^T} = {} & \overline{\underline{c}_j\, \underline{c}_j^T} + 4\gamma_j \left\{ \underline{R}_s\, \overline{\underline{c}_j^T} - \underline{R}_n\, \overline{\underline{c}_j\, \underline{c}_j^T} \right\}^s \\
& - 8\gamma_j^2\, (\underline{R}_s\, \overline{\underline{c}_j^T}\, \underline{R}_n)^s + 4\gamma_j^2\, \underline{R}_s\, \underline{R}_s^T + 4\gamma_j^2\, \overline{z_j^2\, \underline{n}_j\, \underline{n}_j^T}
\end{aligned} \tag{4.68}$$

Similarly, if we take the product of each sides with its transpose in (4.58), we have

$$\underline{c}_{j+1}\,\underline{c}_{j+1}^T = \underline{c}_j\,c_j^T + 2\gamma_j\,(\underline{c}_j\,s_j\,\underline{n}_j^T + s_j\,\underline{n}_j\,\underline{c}_j^T)$$
$$- 2\gamma_j\,z_j\,(\underline{c}_j\,\underline{n}_j^T + \underline{n}_j\,c_j^T)$$
$$- 8\gamma_j^2\,s_j\,z_j\,\underline{n}_j\,\underline{n}_j^T + 4\gamma_j^2\,s_j^2\,\underline{n}_j\,\underline{n}_j^T$$
$$+ 4\gamma_j^2\,z_j^2\,\underline{n}_j\,\underline{n}_j^T \qquad (4.69$$

We shall see that (4.68) and the average of (4.69) are equivalent by virtue of the following terms.

(1) $\overline{\underline{c}_j\,s_j\,\underline{n}_j^T + s_j\,\underline{n}_j\,\underline{c}_j^T} \cong \bar{\underline{c}}_j\,\underline{R}_s^T + \underline{R}_s^T\,\overline{\underline{c}_j}^T$

(2) $\overline{z_j\,(\underline{R}_s\,\underline{n}_j^T + \underline{n}_j\,\underline{R}_s^T)} \cong \overline{z_j\,s_j\,\underline{n}_j\,\underline{n}_j^T} + \overline{z_j\,\underline{n}_j\,s\underline{n}_j^T}$

$$\simeq \overline{2z_j\,s_j\,\underline{n}_j\,\underline{n}_j^T}$$

(3) $\overline{R_s\,R_s^T} = \overline{\overline{s_j\underline{n}_j}\;\overline{s_j\underline{n}_j}}$

$\approx \overline{s_j^2\,n_j\,n_j}$

The last expressions are approximately correct if the number of taps is large. Thus $\overline{c_{j+1}\,c_{j+1}^T}$ either derived from (4.58) or derived from (4.59) are equivalent. Similar steps can be applied to (4.60). In conclusion we can state that the rates of convergence are the same regardless of the choice of algorithms.

V. Adaptive Tapped Delay Line Filters with Time-varying Parameters

The adaptive schemes using the methods of stochastic approximations have been studied for tapped delay line filters. An implicit assumption made so far is that the system under study is time-invariant and all the signals and noise are generated from stationary sources. Although ergodicity of the process has not been required, wide-sense stationarity is implied. If the system itself or the input signals are nonstationary or time-varying, the adjustments made for minimizing certain error criteria may not produce the desired effects. Suppose that the rate of parameter variation is faster than that of convergence, we can never expect to have the algorithms converge at any time. However, if the rate of parameter variation is slow, we can estimate its effects in a qualitative fashion. Let us say that $\varphi(t)$ is a slowly varying time function if the relative change in its value in any interval of length $\Delta t = \frac{2\pi}{\omega_o}$ is small; here ω_o is the minimum frequency of the natural oscillation of the system. If the transient behavior is aperiodic for any initial conditions, the function $\varphi(t)$ is said to be slowly varying when its change is small in comparison with the relative change of the output. The term "slowly varying" used throughout this report is defined in the above sense. The statistical properties of the delay line filter will be studied. For stationary and nonstationary input signals the results seem trivial as a delay element does not change any statistical properties at all, but for the time-varying case the method developed gives us some insights about the system.

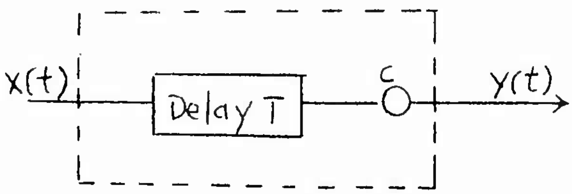
1. Statistical properties of delay line filters[27]

The statistical properties studied here refer only to the auto correlation functions and variance of the output as a measure of the accuracy of the system.

Some other properties like output distributions, probability density functions, etc. are not considered.

A. Stationary case.

Let us first of all consider a single delay element. The input and output are related by



and the transfer function is given by

$$H(j\omega) = \frac{Y(j\omega)}{X(j\omega)} = c\, e^{-j\omega T} \qquad (5.1)$$

If $x(t)$ is a stationary random function with its covariance function function given by

$$R_x(\tau) = D_x\, e^{-\alpha|\tau|} \qquad (5.2)$$

The above expression corresponds to a Markov process and its spectral density function is

$$\Phi_x(\omega) = D_x \int_{-\infty}^{\infty} e^{-\alpha|\tau|} e^{-j\omega T} d\tau$$

$$= D_x \left[\int_{-\infty}^{0} e^{(\alpha\tau - j\omega T)} d\tau + \int_{0}^{\infty} e^{(-\alpha\tau - j\omega T)} d\tau \right]$$

$$= \frac{2\alpha D_x}{\alpha^2 + \omega^2} \qquad (5.3)$$

The output spectral density function is accordingly

$$\Phi_y(\omega) = \Phi_x(\omega) \left| H(j\omega) \right|^2 = \frac{2\alpha D_x}{\alpha^2 + \omega^2} c^2 = \frac{2\alpha c^2 D_x}{\alpha^2 + \omega^2} \qquad (5.4)$$

The output variance is then

$$D_y = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_y(\omega)\, d\omega = \frac{2\alpha c^2}{2\pi} D_x \int_{-\infty}^{\infty}\frac{d\omega}{\alpha^2+\omega^2}$$

$$= D_x \frac{\alpha c^2}{\pi} \cdot 2\pi \cdot \frac{1}{2\alpha} = c^2 D_x \qquad (5.5)$$

as we expected.

A tapped delay line filter consists of several delay elements , gain constants, and a summer so that the transfer function of the filter is

$$H(j\omega) = \sum_{i=0}^{N} c_i e^{-j\omega T i} \qquad (5.6)$$

Direct combination of (5.6), (5.5), and (5.4) yields the output spectrum

$$\phi_y(\omega) = D_x \sum_{i=0}^{N}\sum_{k=0}^{N} c_i c_k e^{-j\omega(T_i-T_k)} \qquad (5.7)$$

and the output variance is the Fourier transform of (5.7)

$$D_y = D_x \sum_{k=0}^{N}\sum_{i=0}^{N} c_k c_i e^{-\alpha|iT - kT|} \qquad (5.8)$$

B. Nonstationary case

Suppose that the input x is a non-stationary time function with the correlation function

$$R_x(t, t') = \sigma_x(t)\,\sigma_x(t')\, e^{-\alpha|t-t'|} \qquad (5.9)$$

The random function x may be expressed by

$$x(t) = \sigma_x(t)\, x_1(t) \qquad (5.10)$$

where $x_1(t)$ is a stationary random time function with correlation function given by (5.2)

$$R_{x1}(\tau) = e^{-\alpha|\tau|} \qquad (5.11)$$

and spectral density function given by (5.3)

$$\phi_{x1}(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2} \tag{5.11}$$

Assuming that the standard deviation $\sigma_x(t)$ of $x(t)$ can be approximated closely enough by the exponential function

$$\sigma_x(t) = \sigma_o e^{\mu t} \tag{5.12}$$

According to 27 , a function $x(t)$ can be put into an integral expansion of the type

$$x(t) = m_x(t) + \sum_{r=1}^{N} \int_{-\infty}^{\infty} v_r(\omega)\, x_r(t, \omega)\, d\omega \tag{5.13}$$

where $m_x(t)$ is the mean value of $x(t)$,

$v_r(\omega)$ are uncorrelated white noise

and

$x_r(t, \omega)$ are coordinate functions defined by

$$x_r(t, \omega) = \sum_{k=1}^{s} C_{rk}(\omega)\, e^{(\mu_k + i\omega)t} \tag{5.14}$$

$r = 1, 2, ---, N$, $C_{rk}(\omega)$ are coefficients.

While the output can also be represented by an integral canonical expansion like $x(t)$ with the coordinate functions

$$Y_r(t, \omega) = \sum_{k=1}^{s} C_{rk}(\omega)\, H(\mu_k + j\omega)\, e^{(\mu_k + j\omega)t} \tag{5.15}$$

$r = 1, 2, ---, N$

The general formulae for the convariance function and dispersion of the output have also been provided by (p. 289, Ref. 27)

$$R_y(t, t') = \sum_{r=1}^{N} \frac{1}{2\pi} \int_{-\infty}^{\infty} G_r(\omega) \sum_{k,\ell=1}^{s} C_{rk}\, C_r^{*}(\omega)$$

$$\cdot\, H(\mu_k + j\omega)\, H^{*}(\mu_\ell + j\omega)\, e^{\mu_k t + \mu_\ell t + j\omega(t-t')} d\omega \tag{5.16}$$

$$D_y(t) = \sum_{r=1}^{N} \frac{1}{2\pi} \int_{-\infty}^{\infty} G_r(\omega) \left| \sum_{k=1}^{s} C_{rk}(\omega) \Phi(\mu_k + j\omega) e^{\mu_k t} \right|^2 d\omega \tag{5.17}$$

where $G_r(\omega)$ are white noise intensities.

In our present case $N = 1$, and the coordinate function is

$$x(t, \omega) = \sigma_o e^{(\mu + j\omega)t} \tag{5.18}$$

and the filter transfer function is modified as

$$H(\mu + j\omega) = \sum_{i=0}^{N} C_i e^{-j(\mu + \omega) T_i} \tag{5.19}$$

Therefore, using (5.16) and (5.17), we have the output correlation function

$$R_y(t, t') = \sum_{i=0}^{N} \frac{\alpha \sigma_o^2 c_i^2}{\pi} e^{(t+t')} \int_{-\infty}^{\infty} \frac{e^{j\omega(t-t')}}{\alpha^2 + \omega^2} d\omega$$

$$= \sigma_o^2 e^{\mu(t+t')} e^{-\alpha|t-t'|} \sum_{i=0}^{N} C_i^2 \tag{5.20}$$

and the variance

$$D_y(t) = \sum_{i=0}^{N} \frac{\alpha D_x(t) C_i^2}{\pi} \int_{-\infty}^{\infty} \frac{d\omega}{\alpha^2 + \omega^2}$$

$$= D_x(t) \sum_{i=0}^{N} C_i^2 \tag{5.21}$$

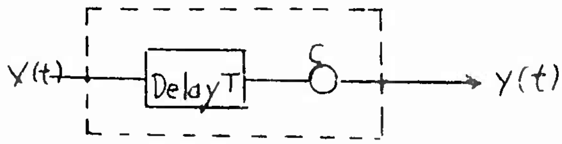
where $D_x(t) = \sigma_x^2(t) = \sigma_o^2 e^{2\mu t}$, the variance of $x(t)$.

C. Time-varying case.

It is a very important case when a linear system is not stationary throughout its total operating time but its behavior is close to it for a comparatively short period. If such a system receives an input which is near to an exponential function, then the output is also near to an exponential function at the end of transient behavior

The technique used here is confined to the situations where only slow time variation is involved.

Consider a single delay element



Since $\qquad y(t) = C\,x(t-T)$

therefore, $\qquad Y(j\omega) = C\,X(j\omega)\,e^{-j\omega T}$

$$e^{+j\omega T}\,Y(j\omega) = c\,X(j\omega)$$

Or in time domain

$$e^{+pT}\,y(t) = c\,x(t) \tag{5.22}$$

where

e^{pT} is an operator represented by

$$e^{pT} = \sum_{j=0}^{\infty} \frac{T^j}{j!}\,\frac{d^j}{dt^j} \quad \text{with } p = \frac{d}{dt} \tag{5.23}$$

At a first glance e^{pT} appears to be a strange-looking operator. Actually $e^{pT}\,y(t) = \sum_{j=0}^{\infty} \frac{T^j}{j!}\,\frac{d^j y(t)}{dt^j}$ is just the Taylor series expansion of $y(t+T)$ around $T=0$. The role of p played here is clear.

Let $x(t) = e^{\lambda t}$ with $\lambda = \mu + j\omega$ and $y(t) = z(t,\lambda)\,e^{\lambda t}$,

then Eq. (5.22) becomes

$$\sum_{j=0}^{\infty} \frac{T^j}{j!}\,\frac{d^j}{dt^j}\left(z(t,\lambda)\,e^{\lambda t}\right) = c\,e^{\lambda t}$$

or

$$z(t,\lambda)\left(e^{pT}\right)e^{\lambda t} + e^{\lambda t}\left(e^{pT}\right)z(t,\lambda) = c\,e^{\lambda t}$$

or

$$z(t,\lambda)\,e^{\lambda T}e^{\lambda t} + e^{\lambda t}\sum_{j=1}^{\infty} \frac{T^j}{j!}\,\frac{\partial^j}{\partial t^j}\left\{z(t,\lambda)\right\} = c\,e^{\lambda t} \tag{5.24}$$

We shall develop a procedure to approximate $z(t,\lambda)$. In the first approximation, we neglect the derivative of the slowly varying function $z(t,\lambda)$ and obtain

$$z_1(t,\lambda)\, e^{\lambda T} = c$$

thus,

$$z_1(t,\lambda) = c\, e^{-\lambda T}$$

and $$y_1(t) = z_1(t,\lambda)\, e^{\lambda t} = c\, e^{\lambda(t-T)} \qquad (5.25)$$

as we expected since $y(t) = c\, x(t-T)$.

In the second approximation, the first derivative of the slowly varying function $z(t,\lambda)$ is taken to be equal to the derivative of $z_1(t,\lambda)$ from the first approximation

$$\frac{\partial z(t,\lambda)}{\partial t} \cong \frac{\partial z_1(t,\lambda)}{\partial t} = \frac{\dot{c}}{e^{\lambda T}} \qquad (5.26)$$

where $\dot{c} = \dfrac{dc(t)}{dt}$

We then get the following equation for the second approximation

$$z_2(t,\lambda)\, e^{\lambda T} + T\,\frac{\partial z_1(t,\lambda)}{\partial t} = c$$

or

$$z_2(t,\lambda)\, e^{\lambda T} = c - T\,\dot{c}\, e^{-\lambda T}$$

$$z_2(t,\lambda) = \frac{1}{e^{\lambda T}}\left(c - T\,\dot{c}\, e^{-\lambda T}\right) \qquad (5.27)$$

the nth approximation of $z(t,\lambda)$ is

$$z_n(t,\lambda) = \frac{1}{e^{\lambda T}}\left(c - \sum_{i=1}^{n} \frac{T^i}{i!}\,\frac{d^i c(t)}{dt^i}\right)$$

depending upon the order n to which the time derivatives of $c(t)$ exist.

For the actual delay line filter, the output is represented by

$$y(t) = \frac{1}{e^{\lambda T}} \sum_{j=0}^{N} \left\{ c_j - \frac{1}{e^{\lambda T}} \sum_{i=1}^{n} \frac{t^i}{i!} \frac{d^i c_j(t)}{dt^i} \right\} \quad (5.28)$$

the variance of $y(t)$ is given respectively by

$$D_y(t) = \frac{\alpha D_x(t)}{\pi} \int_{-\infty}^{\infty} \frac{|z(t,\mu + j\omega)|^2}{\alpha^2 + \omega^2} d\omega \quad (5.29)$$

for the nonstationary input signal used for part B .

If c is time-invariant, then $z(t,\lambda) = c\, e^{-\lambda T}$, (5.29) will be reduced to (5.21) as it should be. The above formulations are only applicable to asymptotically stable systems, and instants which are sufficiently remote from the initial time t .

2. Adaptive schemes for delay line filters with slowly time-varying parameters.

When the system or input characteristics vary slowly with time, it is found convenient to think in terms two time scales by using a "fast" time variable $\tilde{t}$ and a "slow" time variable $\hat{t}$. The ratio of the two scales is a small number so that

$$\hat{t} = \beta \tilde{t} \quad (5.30)$$

The "fast" time variable $\tilde{t}$ refers to the time variable in which the adaptive system operates while the "slow" time variable $\hat{t}$ refers to the time variable in which some parameters in the system vary. One example of the latter case the fluctuation of signal or noise power levels. The levels change but very slowly so that nearly all the techniques developed for time-invariant cases can be applied if additional modifications are made to account for the effect of slow variation. For the adaptive delay line filters under study, if we know the forms of the signal or noise correlation functions (even they are changing very slowly),

the schemes described previously can readily be used. However, if we do not know exactly how the correlation functions change (but we know the cause of variation, for example, sinusoidal or exponential amplitude modulation, frequency modulation, etc.) then we can assume that the weight parameters are functions of slow time variables and leave the unknown fluctuations untouched. In what follows the method of two time variables[28] is described and then applied to the tapped delay line filters with slowly time-varying parameters.

A. Two time variable method

Consider a linear system whose input $x(t)$ and output $e(t)$ are related by the differential equation

$$L e(\tilde{t}, \hat{t}) = M x(\tilde{t}, \hat{t}) \tag{5.31}$$

In the above equation L and M are linear differential operators and can be written in the form

$$L = L(c, \hat{t}, p) = \sum_{i=0}^{n} a_i(c, \hat{t}) p^i \tag{5.32}$$

$$M = M(c, \hat{t}, p) = \sum_{j=0}^{m} b_j(c, \hat{t}) p^j \tag{5.33}$$

where the symbol p denotes the total derivative with respect to time and is defined by

$$p = \frac{d}{dt} = \frac{\partial}{\partial \tilde{t}} + \beta \frac{\partial}{\partial \hat{t}} = \tilde{p} + \beta \hat{p} \tag{5.34}$$

Noting that the adaptive parameter is a slowly varying function of the slow time variable,

$$c = c(\hat{t}) \tag{5.35}$$

We can now expand the operations L and M in Taylor series about $\beta = 0$

$$L(c,\hat{t},p) = L(c,\hat{t},\tilde{p}+\beta\hat{p})$$

$$= L(c,\hat{t},\tilde{p}) + \left.\frac{\partial L}{\partial p}\right|_{\beta=0}(\beta\hat{p}) + \frac{1}{2}\left.\frac{\partial^2 L}{\partial p^2}\right|_{\beta=0}(\beta\hat{p})^2 + \text{----} \tag{5.36}$$

$$M(c,\hat{t},p) = L(c,\hat{t},\tilde{p}+\beta\hat{p})$$

$$= M(c,\hat{t},\tilde{p}) + \left.\frac{\partial M}{\partial p}\right|_{\beta=0}(\beta\hat{p}) + \frac{1}{2}\left.\frac{\partial^2 M}{\partial p^2}\right|_{\beta=0}(\beta\hat{p})^2 + \text{--} \tag{5.37}$$

Since L terminates at the power n and M terminates at the power m,

(5.36) and (5.37) can be rewritten as

$$L = \sum_{j=0}^{n} L_j(\beta\hat{p})^j \tag{5.38}$$

$$M = \sum_{i=0}^{m} M_i(\beta\hat{p})^i \tag{5.39}$$

where

$$L_j = \frac{1}{j!}\left.\frac{\partial^j L}{\partial p^j}\right|_{\beta=0} = \frac{1}{j}\frac{\partial}{\partial\tilde{p}}L_{j-1} = \frac{1}{j!}\frac{\partial^j L_o}{\partial\tilde{p}^j} \tag{5.40}$$

$$M_i = \frac{1}{i!}\left.\frac{\partial^i M}{\partial p^i}\right|_{\beta=0} = \frac{1}{i}\frac{\partial}{\partial\tilde{p}}M_{i-1} = \frac{1}{i!}\frac{\partial^j M_o}{\partial\tilde{p}^j} \tag{5.41}$$

The solution can be expanded in the form

$$e(\hat{t},\check{t}) = \sum_{j=0}^{N}\beta^j e_j \tag{5.42}$$

Substituting the expression of $e(\hat{t},\check{t})$ into (5.31) and equating the terms with same power of β, we have

$$\sum_{j=0}^{n} L_j(\beta\hat{p})^j \sum_{i=0}^{N}\beta^i e_i = \sum_{j=0}^{m}\frac{1}{j!}M_j(\beta\hat{p})^j\sum_{i=0}^{N}\beta^i e_i \tag{5.43}$$

thus, for $j = 0$,

$$L_o\, e_o = M_o\, x \tag{5.44}$$

$j = 1$

$$L_o\, e_1 = -L_1\, \hat{p}\, e_o + M_1\, \hat{p}\, x \tag{5.45}$$

- - - - - - - -

Consequently, $$e_o = \frac{M_o}{L_o}\, x \tag{5.46}$$

$$e_1 = -\frac{1}{L_o}\left(L_1\, \hat{p}\, \left(\frac{M_o}{L_o}\, x\right) - M_1\, \hat{p}\, x\right) \tag{5.47}$$

- - - - - - - - - -

e_o is just the solution for time-invariant case while e_i with $i \geqslant 1$ are additional terms as a result of slow time variation.

The differentiation with respect to $\hat{t}$ implied by $\hat{p}$ can be carried out explicitly.

$$p\left(\frac{M_o}{L_o}\, x\right) = \frac{1}{L_o}\left(\left.\frac{\partial M_o}{\partial \hat{t}}\right|_{c=\text{const}} + \frac{\partial M_o}{\partial c}\, \frac{\partial c(\hat{t})}{\partial \hat{t}}\right) x$$

$$- \frac{M_o}{L_o^2}\left(\left.\frac{\partial L_o}{\partial \hat{t}}\right|_{c=\text{const}} + \frac{\partial L_o}{\partial c}\, \frac{\partial c}{\partial \hat{t}}\right) x$$

$$+ \frac{M_o}{L_o}\, \frac{\partial x}{\partial \hat{t}} \tag{5.48}$$

If the variation in c is slowly enough such that $\frac{\partial^2 c(\hat{t})}{\partial \hat{t}^2} \approx 0$,

then the solution has the form

$$e(\hat{t}, \tilde{t}) \cong e_o + \beta\, e_1 \tag{5.49}$$

where e_1 is obtained by combining (5.47) and (5.48)

$$\begin{aligned} e_1 &= -\frac{1}{L_o}\left(L_1\, \hat{p}\left(\frac{M_o}{L_o}\, x\right) - M_1\, \hat{p}\, x\right) \\ &= -\frac{1}{L_o}\left(L_1\, \hat{p}\, e_o - M_1\, \hat{p}\, x\right) \\ &= -\frac{\partial c(\hat{t})}{\partial \hat{t}}\, \frac{L_1}{L_o}\, \frac{\partial e_o}{\partial c} + \frac{\partial c(\hat{t})}{\partial \hat{t}}\, \frac{M_1}{L_c}\, \frac{\partial x}{\partial c} \end{aligned} \tag{5.50}$$

From Eq. (5.49), the mean square of e is approximately

$$\overline{e^2} = \overline{e_o^{\,2}} + 2\beta\, \overline{e_o e_1} + - - - \qquad (5.51)$$

B. Applications on adaptive filters

Let us now turn our attention to the tapped delay line filter. As has been shown previously, the filter output is

$$z(t) = \sum_{i=0}^{N} c_i \, x(t - T_i) \qquad (5.52)$$

Using Eq. (5.22), we can express Eq. (5.52) as

$$Z(\omega) = \sum_{i=0}^{N} c_i X(\omega) e^{-j\omega T_i} = X(\omega) \sum_{i=0}^{N} c_i \, e^{-j\omega T_i}$$

or in time domain

$$\frac{1}{\sum c_i e^{-p T_i}} \; z(t) = x(t) \qquad (5.53)$$

Comparing with (5.31), we see that

$$L = \frac{1}{\sum_{i=0}^{N} c_i e^{-p T_i}} \qquad (5.54)$$

$M = 1$, unity operator

the operator L_1 is

$$L_1 = \frac{\partial L}{\partial p} = \frac{\partial}{\partial p} \left(\frac{1}{\sum_{i=0}^{N} c_i e^{-p T_i}} \right)$$

$$= \frac{-1}{\left(\sum_{i=0}^{N} c_i e^{-p T_i} \right)^2} \left(\sum_{i=0}^{N} (-T_i)\, c_i e^{-p T_i} \right) \qquad (5.55)$$

and

$$\frac{L_1}{L} = \frac{\sum_{i=0}^{N} T_i c_i e^{-p T_i}}{\sum_{i=0}^{N} c_i e^{-p T_i}} \triangleq T_{av} \qquad (5.56)$$

T_{av} defined by (5.56) can be thought as the average delay time of the filter.

Following (5.49) the filter output is

$$z(t) \cong z_o(\tilde{t}) + \beta\, z_1(\tilde{t}, \hat{t})$$

with

$$z_1(\tilde{t}, \hat{t}) = -\sum_{i=0}^{N} \frac{\partial c_i(\hat{t})}{\partial \hat{t}} \left(\frac{\partial L/\partial p}{L} \right) \frac{\partial z_o(\tilde{t})}{\partial c_i}$$

$$= -\sum_{i=0}^{N} \frac{\partial c_i(\hat{\ })}{\partial \hat{t}} (T_{av}) \frac{\partial}{\partial c_i} \left(\sum_{i=0}^{N} c_i\, x(t - T_i) \right)$$

$$= -T_{av} \sum_{i=0}^{N} \frac{\partial c_i(\hat{t})}{\partial \hat{t}} x(t - T_i) = -T_{av}\, \underline{\eta}^T(t)^T \underline{\delta} \qquad (5.57)$$

where

$$\underline{\eta}(t) = \begin{pmatrix} x(t) \\ x(t-T) \\ \\ x(t-T_N) \end{pmatrix} \quad \text{and} \quad \underline{\delta} = \begin{pmatrix} \delta_0 \\ \delta_1 \\ \\ \delta_N \end{pmatrix} = \begin{pmatrix} \frac{\partial c_o(\hat{t})}{\partial \hat{t}} \\ \vdots \\ \frac{\partial c_N(\hat{t})}{\partial \hat{t}} \end{pmatrix} \qquad (5.58)$$

If the desired signal $d(t)$ is not slowly time-varying, the error function is then

$$e(t) = d(t) - z(t)$$

$$= d(t) - z_o(t) - \beta\, z_i(\tilde{t}, \hat{t})$$

$$= e_o + \beta\, e_1 \qquad (5.59)$$

The mean square error is approximately

$$\overline{e^2(t)} \cong \overline{e_o^2} + 2\beta\,\overline{e_o e_1} \quad . \tag{5.60}$$

Or

$$\overline{e^2(t)} = E\left\{\left(d(t) - \sum_{i=0}^{N} c_i\,\eta_i(t)\right)^2\right\}$$

$$+ 2\beta E\left\{\left(d - \sum_{i=0}^{N} c_i\,\eta_i(t)\right)\left(T\,\underline{\eta}(t)^T\,\underline{\delta}\right)\right\}$$

$$= E\left\{\left(d(t) - \sum_{i=0}^{N} c_i\,\eta_i(t)\right)^2\right\}$$

$$+ 2\beta T_{av} E\left\{\left(d(t) - \sum_{i=0}^{N} c_i\,\eta_i(t)\right)\left(\underline{\eta}^T(t)\right)\underline{\delta}\right\} \tag{5.61}$$

Taking the partial derivative of $\overline{e^2}$ with respect to c_i, we have

$$\frac{\partial\overline{e^2}}{\partial c_i} = 2E\left\{\left(d(t) - \sum_{i=0}^{N} c_i\,\eta_i(t)\right)\left(-\eta_i(t)\right)\right\}$$

$$-2\beta T_{av} E\left\{\eta_i(t)\,\underline{\eta}^T(t)\,\underline{\delta}\right\}$$

Thus the gradient of $\overline{e^2}$ is

$$E\left\{\nabla_c\, Q(\underline{x}\,|\,\underline{c})\right\} = -2E\left\{\underline{\eta}(t)\left(d(t) - \underline{\eta}^T(t)\,c\right)\right\}$$

$$-2\beta T_{av} E\left\{\underline{\eta}(t)\,\underline{\eta}^T(t)\right\}$$

or

$$\nabla_c\, Q(\underline{x}\,|\,\underline{c}) = -2\,\underline{\eta}(t)\left(d(t) - \underline{\eta}^T(t)\,c\right)$$

$$-2\beta T_{av}\,\underline{\eta}(t)\,\underline{\eta}^T(t)\,\underline{\delta} \tag{5.62}$$

Using the adaptive scheme

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \nabla_c Q(\underline{x} \mid \underline{c}) ,$$

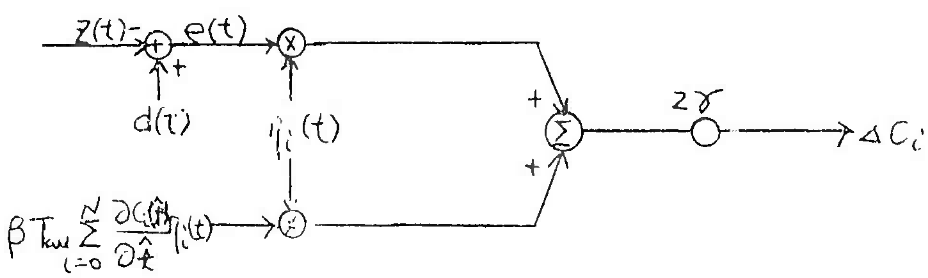
we obtain a new scheme for the time-varying case

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j d_j \underline{\eta}_j - 2\gamma_j \underline{\eta}_j \underline{\eta}_j^T \underline{c}_j$$

$$+ 2\gamma_j \beta T_{av} \underline{\eta}_j \underline{\eta}_j^T \underline{\delta}$$

$$= \underline{c}_j + 2\gamma_j \underline{\eta}_j e_j + 2\gamma_j \beta T_{av} \underline{\eta}_j \underline{\eta}^T \underline{\delta} \qquad (5.63)$$

Algorithm (5.63) can readily be implemented by



Thus the increment for the time-varying parameter c_i is obtained.

Similar schemes can be obtained through proper transformations, Eqs. (4.42) and (4.44), for the cases where only the signal or the noise correlation functions are assumed to be known. The results are

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \beta T_{av} \underline{\eta}_j \underline{\eta}_j^T \underline{\delta}$$

$$- 2\gamma_j \underline{\eta}_j z_j + 2\gamma_j (\underline{\eta}_j x_j - \underline{R}_n) \qquad (5.64)$$

when $R_n(\tau)$ is known, and

$$\underline{c}_{j+1} = \underline{c}_j + 2\gamma_j \beta T_{av} \underline{\eta}_j \underline{\eta}_j^T \underline{\delta}$$

$$- 2\gamma_j \underline{\eta}_j z_j + 2\gamma_j \underline{R}_s \qquad (5.65)$$

when $R_s(\tau)$ is known.

3. Effects of slow time variation on minimum mean square error.

If we set $\overline{\nabla e^2} = 0$, we see from Eq. (5.61) that

$$2\,\overline{\underline{\eta}\;\underline{\eta}^T}\;\underline{c}^* = 2\,\overline{\underline{\eta}\,d} + 2\beta T_{av}\,\overline{\underline{\eta}\;\underline{\eta}^T}\;\underline{\delta}$$

$$\underline{R}_\eta\;\underline{c}^* = \overline{d\,\underline{\eta}} + \beta T_{av}\;\underline{R}_\eta\;\underline{\delta}$$

$$\text{or}\quad \underline{c}^* = \underline{R}_\eta^{-1}\;\overline{d\,\underline{\eta}} + \beta T_{av}\;\underline{\delta} \qquad (5.66)$$

Substituting the expression of $\underline{c}^*$ for $\underline{c}$ into (5.61) and using (3.24 b), we get the minimum mean square error

$$\overline{e^2}_{min} = \overline{d^2} - \overline{d\,\underline{\eta}}^T\;\underline{c}^*$$

$$+\; 2\beta T_{av}\,E\left\{\underline{\delta}^T\underline{\eta}\;(d(t) - \underline{\eta}^T\;\underline{c}^*)\right\}$$

$$\overline{e^2}_{min} = \overline{d^2} - \overline{d\,\underline{\eta}}^T\;\underline{R}_\eta^{-1}\;\overline{d\,\underline{\eta}} - \beta T_{av}\;\overline{d\,\underline{\eta}}^T\;\underline{\delta}$$

$$-\;2\beta T_{av}\,\underline{\delta}^T\;\overline{\underline{\eta}\;\underline{\eta}^T}\;(\underline{R}_\eta^{-1}\;\overline{d\,\underline{\eta}} + \beta T_{av}\;\underline{\delta})$$

$$+\;2\beta T_{av}\,\underline{\delta}^T\;\overline{d\,\underline{\eta}}$$

$$= \overline{d^2} - \overline{d\,\underline{\eta}}^T\;\underline{R}_\eta^{-1}\;\overline{d\,\underline{\eta}} - \beta T_{av}\;\underline{\delta}^T\;\overline{d\,\underline{\eta}}$$

$$-\;2\beta^2 T_{av}^2\;\underline{\delta}^T\;\underline{R}_\eta\;\underline{\delta}$$

$$\leq \overline{e^2_{o\,min}} + \beta T_{av}\left|\underline{\delta}^T\;\overline{d\,\underline{\eta}}\right| + 2\beta^2 T_{av}^2\left|\underline{\delta}^T\;\underline{R}_\eta\;\underline{\delta}\right| \qquad (5.67)$$

where $\overline{e^2_{o\,min}} = \overline{d^2} - \overline{d\,\underline{\eta}}^T\;\underline{R}_\eta^{-1}\;\overline{d\,\underline{\eta}}$ is the minimum mean square error of the time-invariant filter as derived in (3.24 a).

The expressions $\overline{e^2}_{min}$ for other cases (known $R_n(\tau)$) or $R_s(\tau)$) as well as the effect of slow time variation on the rate of convergence can be obtained in a straightforward fashion

Appendix A

Proof of the modified algorithm

It was mentioned in chapter IV that algorithm (4.7) and (4.11) were derived from the formula

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \, (\nabla Q_1 + \overline{\nabla Q_2}) \tag{A.1}$$

rather than from

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \, (\nabla Q_1 + \nabla Q_2) \tag{A.2}$$

where $Q_1 + Q_2 = Q$ is a function of error, and the average of it is the performance criterion to be minimized.

Comparing (A.1) and (A.2) with the regular gradient method with constant γ

$$\underline{c}_{j+1} = \underline{c}_j - \gamma \, (\overrightarrow{\nabla Q_1} + \overline{\nabla Q_2}) \tag{A.3}$$

we see that in (A.2) no average is taken while in (A.1) partial average is taken. The error $\underline{\xi}$ caused by measurement

$$\overline{\nabla Q} = \nabla Q + \underline{\xi}$$

is eliminated by the properly chosen sequence $\{\gamma_j\}$. Intuitively speaking, the same $\{\gamma_j\}$ which eliminates the error caused by $(\overline{\nabla Q_1} + \overline{\nabla Q_2})$ minus $(\nabla Q_1 + \nabla Q_2)$ can definitely eliminate that caused by $(\overline{\nabla Q_1} + \overline{\nabla Q_2})$ minus $(\nabla Q_1 + \overline{\nabla Q_2})$. This stems from the fact that the measuring noise in the second case is smaller on the average than in the first case. Although intuition does not generally warrant mathematical correctness, we can state with mathematical rigour that either signal or noise statistical properties will suffice to generate the error gradient used in the adaptive schemes. The physical conditions under which these algorithms converge remain unchanged. Two lemmas and one theorem will be proved in sequence.

Lemma 1,

For the tapped delay line filters considered in Chapter 4, if

(a) $Q = Q_1 + Q_2$

(b) $Q(e) = e^2$

(c) ∇Q_2 is independent of $\underline{c}$,

then at the neighborhood of $\underline{c}^*$ which minimizes $I(\underline{c}) = E(Q)$, the following statement is true:

$$\inf_{\substack{\varepsilon < \| \underline{c} - \underline{c}^* \| < \frac{1}{\varepsilon} \\ \varepsilon > 0}} E\left\{ (\underline{c} - \underline{c}^*)^T (\nabla Q_1 + \nabla Q_2) \right\} > 0 \tag{A.4}$$

Proof: If $Q = Q_1 + Q_2$ has a minimum at $\underline{c} = \underline{c}^*$, then

$$\frac{\partial(Q_1+Q_2)}{\partial c_i} \begin{cases} > 0 & \text{for} \quad c_i > c_i^* \\ = 0 & \text{for} \quad c_i = c_i^* \\ < 0 & \text{for} \quad c_i < c_i^* \end{cases} \tag{A.5}$$

thus

$$(c_i - c_i^*) \frac{\partial(Q_1 + Q_2)}{\partial c_i} \geq 0 \quad \text{for all } i \tag{A.6}$$

and

$$\inf_{\substack{\varepsilon < \| \underline{c} - \underline{c}^* \| < \frac{1}{\varepsilon} \\ \varepsilon > 0}} E\left\{ (\underline{c} - \underline{c}^*)^T (\nabla Q_1 + \nabla Q_2) \right\} > 0 \tag{A.7}$$

Since ∇Q_2 is independent of $\underline{c}$, we have

$$\begin{aligned} & E\left\{ (\underline{c} - \underline{c}^*)^T (\nabla Q_1 + \nabla Q_2) \right\} \\ = \; & E\left\{ (\underline{c} - \underline{c}^*)^T \nabla Q_1 \right\} + E\left\{ (\underline{c} - \underline{c}^*)^T \nabla Q_2 \right\} \\ = \; & E\left\{ (\underline{c} - \underline{c}^*)^T \nabla Q_1 \right\} + E\left\{ (\underline{c} - \underline{c}^*)^T \right\} E\left\{ \nabla Q_2 \right\} \\ = \; & E\left\{ (\underline{c} - \underline{c}^*)^T (\nabla Q_1 + \nabla Q_2) \right\} \end{aligned} \tag{A.8}$$

Therefore, by virtue of (A.7)

$$\inf_{\substack{\varepsilon < \| \underline{c} - \underline{c}^* \| < \frac{1}{\varepsilon} \\ \varepsilon > 0}} E\left\{ (\underline{c} - \underline{c}^*)^T (\nabla Q_1 + \nabla Q_2) \right\} > 0 \tag{A.9}$$

Lemma 2

If $Q = Q_1 + Q_2$, $Q(e) = e^2$, ∇Q_2 is independent of $\underline{c}$, and

(a) $\frac{\partial^2 Q_1}{\partial e^2}$ exists and is uniformly bounded

(b) $s(t)$ and $n(t)$ are uniformly bounded,

then for the tapped delay line filter

$$E\left\{(\nabla Q_1 + \overline{\nabla Q_2})^T \quad (\nabla Q_1 + \overline{\nabla Q_2})\right\} \leq d(\underline{c}^{*T}\underline{c}^* + \underline{c}^T\underline{c}) \qquad d > 0 \tag{A.10}$$

Proof:

Using a Taylor expansion about $\underline{c} = \underline{c}^*$, we have

$$\frac{\partial Q_1}{\partial c_j} = \left.\frac{\partial Q_1}{\partial c_j}\right|_{\underline{c} = \underline{c}^*} + \sum_{i=0}^{N} (c_i - c_i^*) \left.\frac{\partial^2 Q_1}{\partial c_i \partial c_j}\right|_{\underline{c} = \underline{c}^*} \tag{A.11}$$

for arbitrary j , and $j = 0$, 1, 2, ---, N .

Since $e(t) = s(t) - z(t)$

$$= s(t) - \sum_{i=0}^{N} c_i n_i(t) \tag{A.12}$$

we see that

$$\frac{\partial Q_1(e)}{\partial c_i} = \frac{\partial Q_1}{\partial e} \frac{\partial e}{\partial c_i} = \frac{\partial Q_1}{\partial e} [- n_i(t)] \tag{A.13}$$

and

$$\frac{\partial^2 Q_1(e)}{\partial c_i \partial c_j} = \frac{\partial^2 Q_1}{\partial e^2} n_i(t) \quad n_j(t)$$

$$= \frac{\partial^2 Q_1}{\partial e^2} [s(t - T_i) + n(t - T_i)] [s(t - T_j) + n(t - T_j)] \tag{A.14}$$

Thus

$$\frac{\partial^2 Q_1}{\partial c_i \partial c_j} \quad \text{is bounded if (a) and (b) are satisfied.}$$

Therefore, from (A.11)

$$\nabla_c Q_1 \leq \nabla_c Q_1 \Big|_{c = c^*} + k_1 \sum_{i=0}^{N} (c_i - c_i^*) \tag{A.15}$$

$$\text{where} \quad k_1 = k \sup_{\text{all } i} \left| \frac{\partial^2 Q_1}{\partial c_i \partial c_j} \right| \tag{A.16}$$

As Q_2 contains $\underline{c}$ only in the first order and $\nabla_c Q_2$ is independent of $\underline{c}$, we can write

$$\overline{\nabla_c Q_2} = \overline{\nabla_c Q_2} \Big|_{c = c^*} \tag{A.17}$$

and

$$\nabla Q_1 + \overline{\nabla Q_2} \leq \nabla Q_1 \Big|_{c = c^*} + \overline{\nabla Q_2} \Big|_{c = c^*} + k_1 \sum_{i=0}^{N} (c_i - c_i^*) \tag{A.18}$$

Note

$$E\left\{ \nabla Q_1 \Big|_{c = c^*} + \overline{\nabla Q_2} \right\} = E\left\{ \nabla Q_1 + \overline{\nabla Q_2} \right\} \Big|_{c = c^*}$$

$$= E\left\{ \nabla Q_1 + \nabla Q_2 \right\} \Big|_{c = c^*} = 0 \tag{A.19}$$

Taking mathematical expectation on both sides of (A.17) gives

$$E\left\{ \nabla Q_1 + \overline{\nabla Q_2} \right\} \leq k_1 \sum_{i=0}^{N} (c_i - c_i^*) \tag{A.20}$$

Lemma 2 is obtained by taking the inner product of (A.20)

$$E\left\{ (\nabla Q_1 + \overline{\nabla Q_2}) (\nabla Q_1 + \overline{\nabla Q_2}) \right\}$$

$$\leq k_1^2 \sum_{i=0}^{N} \sum_{j=0}^{N} (c_i - c_i^*) (c_j - c_j^*)$$

$$= d(\underline{c}^{*T} \underline{c}^* + \underline{c}^T \underline{c}) \tag{A.21}$$

Theorem

Let γ_1 , γ_2 - - - be a sequence of positive numbers such that

$$\text{(A1)} \quad \lim_{j\to\infty} \gamma_j = 0$$

$$\text{(A2)} \quad \sum_{j=0}^{\infty} \gamma_j = \infty \qquad \text{(A.23)}$$

$$\text{(A3)} \quad \sum_{j=1}^{\infty} \gamma_j^2 < \infty$$

Let the following conditions be satisfied

$$\text{(B)} \quad \inf_{\varepsilon < ||c - c^*|| < \frac{1}{\varepsilon}} E \left\{ (c - c^*)^T (\nabla_c Q_1 + \overline{\nabla Q_2}) \right\} > 0 \qquad \text{(4.24)}$$

$$,\varepsilon > 0$$

$$\text{(C)} \quad E \left\{ (\nabla_c Q_1 + \overline{\nabla Q_2})^T (\nabla Q_1 + \overline{\nabla Q_2}) \right\}$$

$$< d \, (\underline{c}^{*T} \underline{c}^* + \underline{c}^T \underline{c}) \quad , \qquad \text{(A.25)}$$

$d > 0$, for all c in a bounded set.

Then the algorithm

$$\underline{c}_{j+1} = \underline{c}_j - \gamma_j \, (\nabla Q_1 + \overline{\nabla Q_2}) \qquad \text{(A.26)}$$

which minimizes the performance criterion

$$I(c) = E \left\{ Q(e) \right\} = E \left\{ Q_1(e) + Q_2(e) \right\} \qquad \text{(A.27)}$$

converges with probability and to $\underline{c}^*$.

Proof:

Substrating both sides of (A.26) by $\underline{c}^*$, we have

$$\underline{c}_{j+1} - \underline{c}^* = \underline{c}_j - \underline{c}^* - \gamma_j \, (\nabla Q_1 + \overline{\nabla Q_2}) \qquad \text{(A.28)}$$

taking the inner product on both sides of (A.23)

$$(\underline{c}_{j+1} - \underline{c}^*)^T (\underline{c}_{j+1} - \underline{c}^*)$$

$$= (\underline{c}_j - \underline{c}^*)^T (\underline{c}_j - \underline{c}^*) - 2 \gamma_j (c_j - \underline{c}^*)^T (\nabla Q_1 + \overline{\nabla Q_2})$$

$$+ \gamma_j^2 (\nabla Q_1 + \overline{\nabla Q_2})^T (\nabla Q_1 + \overline{\nabla Q_2}) \qquad \text{(A.29)}$$

and taking the conditional mathematical expectation for given $\underline{c}_1$, $\underline{c}_2$, ... , $\underline{c}_j$, we obtain

$$E\left\{ ||\underline{c}_{j+1} - \underline{c}^*||^2 \,\middle|\, \underline{c}_1 , \underline{c}_2 , ---, \underline{c}_j \right\}$$

$$= ||\underline{c}_j - \underline{c}^*||^2 - 2\gamma_j \; E\left\{ (\underline{c}_j - \underline{c}^*)^T \, (\nabla Q_1 + \overline{\nabla Q}_2) \right\}$$

$$+ \gamma_j^{\,2} \; E\left\{ (\nabla Q_1 + \overline{\nabla Q}_2)^T \; (\nabla Q_1 + \overline{\nabla Q}_2) \right\} \tag{A.30}$$

From condition (c), (A.30) becomes

$$E\left\{ ||\underline{c}_{j+1} - \underline{c}^*||^2 \,\middle|\, \underline{c}_1, --- \underline{c}_j \right\}$$

$$< ||\underline{c}_j - \underline{c}^*||^2 - 2\gamma_j \; E\left\{ (\underline{c}_j - \underline{c}^*)^T \, (\nabla Q_1 + \overline{\nabla Q}_2) \right\}$$

$$+ \gamma_j^{\,2} \; d(c^{*T} \, c^* + c_j^{\,T} \, c_j) \tag{A.31}$$

Using condition (B), (A.31) is reduced to

$$E\left\{ ||\underline{c}_{j+1} - \underline{c}^*|| \,\middle|\, \underline{c}_1, -- \underline{c}_n \right\}$$

$$< ||\underline{c}_j - \underline{c}^*||^{\cdot} \; (1 + 2\gamma_j^{\,2} \, d) + 2\gamma_j^{\,2} \, d \, \underline{c}^T \, \underline{c}^* \tag{A.31a}$$

Using condition (B), we can reduce (A.31a) to

$$E\left\{ ||\underline{c}_{j+1} - \underline{c}^*||^2 \,\middle|\, \underline{c}_1, ---, \underline{c}_n \right\}$$

$$< ||\underline{c}_j - \underline{c}^*||^2 \; (1 + 2\gamma_j^{\,2} \, d) + 2\gamma_j^{\,2} \, d \, \underline{c}^T \, \underline{c}^* \tag{A.32}$$

Let

$$\underline{z}_j = ||\underline{c}_j - \underline{c}^*||^2 \prod_{k=j}^{\infty} (1 + \gamma_k^{\,2} \, d)$$

$$+ \sum_{k=j}^{\infty} 2d \, \gamma_k^{\,2} \, \underline{c}^T \, \underline{c}^* \prod_{m=k+1}^{\infty} (1 + \gamma_m^{\,2} \, d) \tag{A.33}$$

Then

$$\underline{z}_{j+1} = ||\underline{c}_{j+1} - \underline{c}^*||^2 \prod_{k=j+1}^{\infty} (1 + \gamma_k^{\,2} \, d)$$

$$+ \sum_{k=j+1}^{\infty} 2 \, d \, \gamma_k^{\,2} \, \underline{c}^T \, \underline{c}^* \prod_{m=k+1}^{\infty} (1 + \gamma_m^{\,2} \, d) \tag{A.34}$$

Taking the conditional mathematical expectation for given $\underline{c}_1, \underline{c}_2, \text{---} \underline{c}_j$,

we have

$$E\left\{ \underline{Z}_{j+1} \middle| c_1, \text{--} c_n \right\} = E\left\{ ||c_{j+1} - c^*||^2 \middle| c_1, \text{--} c_n \right\} \prod_{k=j+1}^{\infty} (1 + \gamma_k^2 d)$$

$$+ \sum_{k=j+1}^{\infty} 2\, d\, \gamma_k^2\, c^T c^* \prod_{m=k+1}^{\infty} (1 + \gamma_m^2 d)$$

$$\leq \left[||c_j - c^*||^2 (1 + d\, \gamma_j^2) + 2\, \gamma_j^2\, d\, c^T c^* \right] \prod_{k=j+1}^{\infty} (1 + \gamma_k^2 d)$$

$$+ \sum_{k=j+1}^{\infty} 2 d \gamma_k^2\, c^T c^* \prod_{m=k+1}^{\infty} (1 + \gamma_m^2 d)$$

$$= \underline{Z}_j$$

or

$$E\left\{ \underline{Z}_{j+1} \middle| c_1, \text{--}, c_n \right\} \leq \underline{Z}_j \tag{A.35}$$

Next taking the conditional mathematical expectation for given $\underline{Z}_1, \text{---}, \underline{Z}_j$ on both sides of (A.35) we have

$$E\left\{ \underline{Z}_{j+1} \middle| Z_1, \text{---}, Z_j \right\} \leq \underline{Z}_j \tag{A.36}$$

Inequality (A.36) shows that Z_j is a semimartingale, where

$$E\ \underline{Z}_{j+1} < E\ \underline{Z}_j < \text{---} < E\ \underline{Z}_1 < \infty \tag{A.37}$$

so that, according to the theory of semimartingales[22] the sequence Z_1 converges with probability one, and hence by virtue of (A.33) and (A.23c) the sequence $(\underline{c}_j - \underline{c}^*)$ also converges with probability one to some random number ξ . It remains to show that $P(\underline{\xi} = 0) = 1$.

It is seen that from (A.37), (A.33) and (A.23c) the sequence $E(c_j - c^*)$ is bounded. Now taking the mathematical expectation on both side of the inequality (A.32),

$$E\left\{ ||\underline{c}_{j+1} - \underline{c}^*||^2 \right\} < E\left\{ ||\underline{c}_j - \underline{c}^*||^2 \right\} - 2\gamma_j\, E\left\{ (\underline{c}_j - \underline{c}^*)^T \nabla Q \right\}$$

$$+ \gamma_j^2\, d\, [c^{*T} c^* + E(c_j^T c_j)]$$

and adding the first j inequalities together, we have by deduction

$$E\left\{||\underline{c}_{j+1} - \underline{c}^*||^2\right\} \leq E\left\{||c_1 - c^*||^2\right\} + \sum_{k=1}^{j} [\underline{c}^{*T} \underline{c}^* \gamma_k^2 + d\, \gamma_k^2\, E(\underline{c}_j^T \underline{c})]$$

$$- \sum_{k=1}^{j} 2\, \gamma_k\, E\left\{(c_j - c^*)^T \nabla Q\right\} \tag{A.38}$$

Since $E\left\{||c_j - c^*||^2\right\}$ is bounded and condition (A.23c) is fulfilled, from Eq. (A.38) it follows that

$$\sum_{k=1}^{\infty} \gamma_k\, E\left\{(c_j - c^*)^T \nabla Q\right\} < \infty \tag{A.39}$$

Using condition (A.23b), i.e., $\sum_{j=1}^{\infty} \gamma_j = \infty$

and noting (A.24)

$$\inf_{\varepsilon < ||\underline{c} - \underline{c}^*|| < \frac{1}{\varepsilon}} E\left\{(\underline{c} - \underline{c}^*)^T \nabla Q\right\} > 0$$

We deduce from (A.39) that

$$E\left\{(\underline{c}_N - \underline{c}^*)^T \nabla Q\right\} \to 0 \quad \text{with probability one for some sequence } N \tag{A.40}$$

Now taking $E\left\{||\underline{c}_j - \underline{c}^*||^2\right\} \to \underline{\xi}$ with probability 1, and comparing (A.40) with (A.24), we obtain

$$\underline{\xi} = 0 \quad \text{with probability 1} \tag{A.41}$$

Therefore, algorithm (A.26) converges with probability one

$$P\left\{\lim_{j\to\infty} (\underline{c}_j - \underline{c}^*) = 0\right\} = 1 \tag{A.42}$$

as well as in mean square sense, i.e.,

$$\lim_{j\to\infty} E\left\{||\underline{c}_j - \underline{c}^*||^2\right\} = 0 \tag{A.43}$$

Appendix B

Some properties of Gamma functions

Since $\Gamma(\alpha + n) = (\alpha + n - 1)\,\Gamma(\alpha + n - 1)$

$$= (\alpha + n - 1)(\alpha + n - 2)\Gamma(\alpha + n - 2)$$

$$= - - - -$$

$$= (\alpha + n - 1)(\alpha + n - 2) --- \alpha\Gamma(\alpha)$$

We have

$$\prod_{k=1}^{n} (\alpha + k - 1) = \alpha(\alpha + 1) --- (\alpha + n - 1)$$

$$= \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)} \tag{B.1}$$

Thus Eq. (4.29) becomes

$$\prod_{k=1}^{j} \left(1 - \frac{\lambda}{j+1}\right) = \frac{\prod_{k=1}^{j} (j+1 - \lambda)}{(j+1)!} = \frac{\Gamma(j+2 - \lambda)}{(j+1)!\,\Gamma(2 - \lambda)} \tag{B.2}$$

Eq.(B.2) can be approximated by using the formula*

$$\Gamma(x) = e^{-x}\, x^{x-\frac{1}{2}}\, (2\pi)^{\frac{1}{2}} \left\{ 1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320\,x^4} + 0\left(\frac{1}{x5}\right) \right\}$$

$$\cong e^{-x}\, x^{x-\frac{1}{2}}\, (2\pi)^{\frac{1}{2}} \quad \text{for} \quad x >> 1 \tag{B.3}$$

From Eq. (B.3) we can write for $j >> 1$,

$$\Gamma(j + 2 - \alpha) \simeq e^{-(j+2-\alpha)}\, (j+2-\alpha)^{j+2-\alpha-\frac{1}{2}}\, (2\pi)^{\frac{1}{2}}$$

$$= e^{-(j+2-\alpha)}\, (j+2-\alpha)^{j+\frac{3}{2}}\, (j+2-\alpha)^{-\alpha}\, (2\pi)^{\frac{1}{2}} \tag{B.4}$$

* Whittaker and Waston, Modern Analysis, p. 253

$$(j+1)! = \Gamma(j+2) \cong e^{-(j+2)} (j+2)^{j+\frac{3}{2}} (2\pi)^{\frac{1}{2}} \tag{B.5}$$

Since

$$j + 2 - \alpha \cong j+2 \quad \text{if} \quad j >> \alpha$$

we obtain from (B.4) and (B.5)

$$\frac{\Gamma(j+2-\alpha)}{(j+1)!} = \frac{\Gamma(j+2-\alpha)}{\Gamma(j+2)} \cong (j+2-\alpha)^{-\alpha}$$

$$\cong \frac{1}{(j+1)^{\alpha}} \quad \text{if} \quad j >> 1 \quad \text{and} \quad j >> \alpha \tag{B.6}$$

Therefore, combining (B.2) and (B.6) gives

$$\prod_{k=1}^{j} (1 - \frac{\lambda}{j+1}) \cong \frac{1}{\Gamma(2-\lambda)(j+1)^{\lambda}} \tag{B.7}$$

and furthermore,

$$\prod_{j=1}^{m} (1 - \frac{A}{j+1}) \cong \frac{m^A}{(n+1)^A} \tag{B.8}$$

REFERENCES

1. Wiener, N., "Extrapolation, Interpolation and Smoothing of Stationary Time Series," Wiley, New York, 1949.

2. Davenport, W. B., and Root, W. L., "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Co., New York, 1958.

3. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," J. Basic Engineering, Series D, Vol. 83, 95-108, March 1961.

4. Bucy, R. S., and Follin, J. W., "Adaptive Finite Time Filtering," IRE Trans. on Automatic Control, Vol. AC-7, 10-19, July 1962.

5. Kalman, R. E., "Design of a Self-Optimizing Control System," ASME Trans., Vol. 80, No. 2, 468-478, February 1958.

6. Widrow, B., "Adaptive Sampled-Data Systems - A Statistical Theory of Adaptation," 1959 WESCON Conv. Rec., IRE, Part 4, 1959.

7. Narendra, K. S., and McBride, L., "Multiparameter Self-Optimizing Systems Using Correlation Techniques," IEEE Trans. PGAC, AC-9, 31-38, January 1964.

8. McGrath, R. J., and Rideout, V. C., "A Simulator Study of a Two-Parameter Adaptive System," IEEE Trans. PGAC, AC-6, 35-42, February 1961.

9. Gabor, D., etc., "A Universal Non-Linear Filter, Predictor and Simulator Which Optimizes Itself by a Learning Process," Proc. of IEE, Paper No. 3270M, 422-438, July 1960.

10. Weaver, C. S., "Adaptive Communication Filtering," IRE Trans. PGIT, Vol. IT-8, 169-178, September 1962.

11. Narendra, K. S. Galman, P., and Chang, J. H., "Identification of Non-linear Systems Using Gradient and Iterative Techniques," Dunham Lab Technical Report CT-3, Yale University, August 1966.

12. Widrow, B., "Adaptive Sampled-Data Systems," Proc. First International Congress of the International Federation of Automatic Control," Moscow 1960.

13. Widrow, B., "Adaptive Filters 1: Fundamentals," Technical Report No. 6764-6, System Theory Lab. Stanford University, December 1966.

14. Robbins, H., and Monro, S., "A Stochastic Approximation Method," Am. Math. Statistic, Vol. 22, No. 1, 400-407, 1951.

15. Kiefer, J., and Wolfowitz, J., "Stochastic Estimation of the Maximum of a Regression Function." Am. Math. Stat., Vol. 23, 462-466, 1952.

16. Blum, J. R., "Multidimensional Stochastic Approximation Methods," Am. Math. Stat., Vol. 25, 737-744, 1954.

17. Dvoretzky, A., "On Stochastic Approsimation," Proc. 3d Berkeley Sym. on Math. Stat. and Prob., 39-55, 1956.

18. Kesten, H., "Accelerated Stochastic Approximation," Am. Math. Stat. Vol. 29, 41-59, 1958.

19. Wilde, D. J., "Optimum Seeking Methods," Prentice Hall, 1964.

20. Tsypkin, Ya. Z., "Adaptation, Learning and Self-learning in Automatic Systems," Automation and Remote Control, Vol 27, January 1966.

21. Gladysnev, E. G., "On Stochastic Approximation," Theory of Probability and its Applications, Vol. 10, No. 2, 275-278, 1965.

22. Doob, J. L., "Stochastic Processes," John Wiley and Sons, New York 1953.

23. Chang, S. S. L., "Synthesis of Optimum Control Systems," McGraw-Hill Book Co., 1961.

24. Weinberg, L., "Network Analysis and Synthesis," McGraw-Hill, 1962.

25. Knapp, C. H., "Optimum Linear Filtering for Multi-Element Arrays," T. R. U417-66-031, Electric Boat Div., General Dynamics Corp., Nov. 1966.

26. Sabrison, D. T., "Application of Stochastic Approximation Methods to Optimal Filter Design," I E International Convention Record, Vol. 9, Part 4, 1961.

27. Pugachev, V. S., "Theory of Random Functions," English Translation, Pergamon Press, N. Y. 1965.

28. Kronaur, R. E., and Drew, P. G., "Design of the Adaptive Feedback Loop in the Parameter-Perturbation Adaptive Controls," presented at the International Federation on Automatic Control (Teddington) Symposium 1965 on the Theory of Self-Adaptive Control Systems, Teddington, England, September 1965.

29. Comer, T. R., "Some Stochastic Approximation Procedures for Use in Process Control," Amer. Math. Stat., Vol. 35, No. 3, 1964.